

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 2004年 1月16日
Date of Application:

出願番号 特願2004-009004
Application Number:
[ST. 10/C]: [JP 2004-009004]

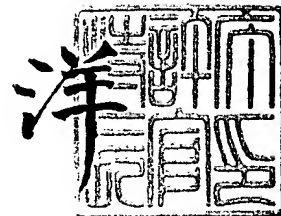
出願人 株式会社日立製作所
Applicant(s):

CERTIFIED COPY OF
PRIORITY DOCUMENT

2004年 7月 2日

特許庁長官
Commissioner,
Japan Patent Office

小川



出証番号 出証特2004-3057551

【書類名】 特許願
【整理番号】 340301147
【提出日】 平成16年 1月16日
【あて先】 特許庁長官殿
【国際特許分類】 G06F 3/06
【発明者】
 【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A
 I D システム事業部内
 【氏名】 山崎 孝
【発明者】
 【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A
 I D システム事業部内
 【氏名】 袴田 和夫
【発明者】
 【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A
 I D システム事業部内
 【氏名】 加納 東
【特許出願人】
 【識別番号】 000005108
 【氏名又は名称】 株式会社日立製作所
【代理人】
 【識別番号】 110000176
 【氏名又は名称】 一色国際特許業務法人
 【代表者】 一色 健輔
【手数料の表示】
 【予納台帳番号】 211868
 【納付金額】 21,000円
【提出物件の目録】
 【物件名】 特許請求の範囲 1
 【物件名】 明細書 1
 【物件名】 図面 1
 【物件名】 要約書 1

【書類名】 特許請求の範囲**【請求項 1】**

情報処理装置と通信可能に接続され、

複数のハードディスクドライブと、

前記情報処理装置から前記ハードディスクドライブに対するデータの読み出し要求と書き込み要求とを受信するホストインタフェースと、通信経路を介して前記ハードディスクドライブと通信可能に接続され前記ハードディスクドライブとの間でデータの入出力を行うディスクインタフェースと、メモリと、前記ホストインタフェースと前記ディスクインタフェースとの制御を司るCPUと、計時機構とを含んで構成されるコントローラと

を有し、

複数の前記ハードディスクドライブにより構成される冗長性を有するRAIDグループに論理ボリュームが形成され、

前記情報処理装置から前記論理ボリュームに対するデータの読み出し要求又は書き込み要求を受信すると前記計時機構より取得される時刻を前記メモリにアクセス時刻として当該論理ボリュームが形成されている前記RAIDグループの識別子と対応付けて記憶するアクセス時刻記憶部と、

前記メモリに記憶されている前記アクセス時刻を参照し、前記計時機構より取得される時刻と前記アクセス時刻との差が所定の時間を超えている場合、前記RAIDグループの冗長性に応じた数の前記ハードディスクドライブを節電モードとする節電モード実行部とを有することを特徴とするディスクアレイ装置。

【請求項 2】

請求項 1 に記載のディスクアレイ装置であって、

前記ディスクインタフェースと前記複数のハードディスクドライブとを接続する複数の通信経路を備え、

前記ハードディスクドライブは前記通信経路のうちのいずれかと通信可能に接続され、

節電モードとなっている前記ハードディスクドライブの数が最も少ない前記通信経路を選択する通信経路選択部を有し、

前記節電モード実行部は、前記通信経路選択部により選択される前記通信経路と通信可能に接続されている前記ハードディスクドライブを節電モードとすること

を特徴とするディスクアレイ装置。

【請求項 3】

請求項 1 に記載のディスクアレイ装置であって、

前記情報処理装置から前記節電モードとなっている前記ハードディスクドライブを含んで形成される前記論理ボリュームに対する前記読み出し要求を受信すると、前記論理ボリュームを構成する前記ハードディスクドライブのうちの節電モードとなっていない前記ハードディスクドライブから前記RAIDグループの冗長性を用いて前記読み出し要求に回答する読み出し要求応答部と、

前記読み出し要求応答部が前記読み出し要求に回答した後に、節電モードとなっている前記ハードディスクドライブの節電モードを解除する節電モード解除部と

を有することを特徴とするディスクアレイ装置。

【請求項 4】

請求項 1 に記載のディスクアレイ装置であって、

前記節電モード実行部が前記ハードディスクドライブを節電モードとする前に、当該ハードディスクドライブに記憶されているデータの複製を予備の前記ハードディスクドライブであるスペアドライブに記憶するスペアドライブ記憶部と、

節電モードとなっている前記ハードディスクドライブを含んで形成される前記論理ボリュームに対する前記書き込み要求を受信すると、前記論理ボリュームを構成する前記ハードディスクドライブのうちの節電モードとなっていない前記ハードディスクドライブと前記スペアドライブとを前記論理ボリュームが形成されている前記RAIDグループとして前記書き込み要求に回答する書き込み要求応答部と、

前記書き込み要求応答部が前記書き込み要求に応答した後に、節電モードとなっている前記ハードディスクドライブの節電モードを解除し、前記スベアドライブに記憶されているデータの複製を節電モードが解除された前記ハードディスクドライブに記憶する節電モード解除部と

を有することを特徴とするディスクアレイ装置。

【請求項 5】

請求項 4 に記載のディスクアレイ装置であって、

前記書き込み要求応答部が前記書き込み要求に応じて前記スベアドライブに書き込んだデータの位置情報を前記メモリに記憶する位置情報記憶部を有し、

前記節電モード解除部は、前記スベアドライブの前記位置情報で示される位置に記憶されている前記データの複製を節電モードが解除された前記ハードディスクドライブに記憶すること

を特徴とするディスクアレイ装置。

【請求項 6】

請求項 1 に記載のディスクアレイ装置であって、

節電モードとなっている前記ハードディスクドライブを含んで形成される前記論理ボリュームに対する前記書き込み要求を受信すると、当該書き込み要求に伴うデータを前記論理ボリュームを構成する前記ハードディスクドライブのうちの節電モードとなっていない前記ハードディスクドライブのみに書き込んで当該書き込み要求に応答する書き込み要求応答部と、

前記書き込み要求応答部が前記書き込み要求に応答した後に、節電モードとなっている前記ハードディスクドライブの節電モードを解除し、前記論理ボリュームを構成する前記ハードディスクドライブのうちの節電モードとなっていなかった前記ハードディスクドライブに記憶されているデータから前記 R A I D グループの冗長性を用いて節電モードとなっていた前記ハードディスクドライブに記憶されるデータを生成し、生成された当該データを節電モードが解除された前記ハードディスクドライブに記憶する節電モード解除部とを有することを特徴とするディスクアレイ装置。

【請求項 7】

請求項 6 に記載のディスクアレイ装置であって、

前記書き込み要求において節電モードとなっている前記ハードディスクドライブに書き込まれるデータの当該ハードディスクドライブにおける位置情報を前記メモリに記憶する位置情報記憶部を有し、

前記節電モード解除部は、前記論理ボリュームを構成する前記ハードディスクドライブのうちの節電モードとなっていなかった前記ハードディスクドライブに記憶されているデータから前記 R A I D グループの冗長性を用いて節電モードとなっていた前記ハードディスクドライブの前記位置情報で示される位置に記憶されるデータを生成し、生成された当該データを節電モードが解除された前記ハードディスクドライブに記憶すること

を特徴とするディスクアレイ装置。

【請求項 8】

情報処理装置と通信可能に接続され、

第一のインタフェース規格によりデータの送受信を行う複数の第一のハードディスクドライブと、

前記第一のハードディスクドライブよりも寿命の短い、第二のインタフェース規格によりデータの送受信を行う複数の第二のハードディスクドライブと、

前記情報処理装置から前記第一又は第二のハードディスクドライブに対するデータの読み出し要求と書き込み要求とを受信するホストインタフェースと、通信経路を介して前記第一又は第二のハードディスクドライブと通信可能に接続され前記第一又は第二のハードディスクドライブとの間でデータの入出力を行うディスクインタフェースと、メモリと、前記ホストインタフェースと前記ディスクインタフェースとの制御を司る C P U と、計時機構とを含んで構成されるコントローラと

を有し、

複数の前記第一又は第二のハードディスクドライブにより構成される R A I D グループに論理ボリュームが形成され、

前記情報処理装置から前記論理ボリュームに対するデータの読み出し要求又は書き込み要求を受信すると前記計時機構より取得される時刻をアクセス時刻として当該論理ボリュームが形成されている前記 R A I D グループの識別子と対応付けて前記メモリに記憶するアクセス時刻記憶部と、

前記メモリに記憶されている前記アクセス時刻を参照し、前記計時機構より取得される時刻と前記アクセス時刻との差が所定の時間を超えている前記 R A I D グループについて、前記 R A I D グループが前記第一のハードディスクドライブのみで構成されている場合は、前記 R A I D グループの冗長性に応じた数の前記第一のハードディスクドライブを節電モードとし、前記 R A I D グループが前記第二のハードディスクドライブのみで構成されている場合は、任意の数の前記第二のハードディスクドライブを節電モードとする節電モード実行部と、

前記節電モード実行部により前記第一又は第二のハードディスクドライブが節電モードとなると、前記計時機構より取得される時刻を節電開始時刻として前記 R A I D グループの識別子と対応付けて前記メモリに記憶する節電開始時刻記憶部と、

前記節電開始時刻を参照し、当該節電開始時刻と前記計時機構より取得される時刻との差が所定の時間を超えている場合は、節電モードとなっている前記第一又は第二のハードディスクドライブの節電モードを解除する節電モード解除部と

を有することを特徴とするディスクアレイ装置。

【請求項 9】

請求項 8 に記載のディスクアレイ装置であって、

前記 R A I D グループ毎の利用形態が前記メモリに記憶され、

前記節電モード実行部は、前記 R A I D グループが前記第一のハードディスクドライブと前記第二のハードディスクドライブとで構成されている場合は、前記 R A I D グループの前記利用形態に応じて、前記 R A I D グループの冗長性に応じた数又は任意の数の前記第一又は第二のハードディスクドライブを節電モードとすること

を特徴とするディスクアレイ装置。

【請求項 10】

請求項 8 に記載のディスクアレイ装置であって、

前記 R A I D グループ毎の連続稼働時間を前記メモリに記憶する連続稼働時間記憶部と

、
前記連続稼働時間が所定の時間を超えている前記 R A I D グループを構成する全ての前記第一又は第二のハードディスクドライブに記憶されているデータの複製を予備の前記第一又は第二のハードディスクドライブであるスペアドライブに記憶し、前記 R A I D グループを構成する全ての前記第一又は第二のハードディスクドライブを節電モードとする一括スペア実行部と

を有することを特徴とするディスクアレイ装置。

【請求項 11】

請求項 8 に記載のディスクアレイ装置であって、

前記第一又は第二のハードディスクドライブ毎の累積稼働時間を前記メモリに記憶する累積稼働時間記憶部を有し、

前記節電モード実行部は、前記 R A I D グループを構成する前記第一又は第二のハードディスクドライブのうちで前記累積稼働時間が長い前記第一又は第二のハードディスクドライブを節電モードとすること

を特徴とするディスクアレイ装置。

【請求項 12】

請求項 11 に記載のディスクアレイ装置であって、

節電モード解除部が節電モードとなっている前記第一又は第二のハードディスクドライブ

ブの節電モードを解除した後に、前記節電モード実行部は、当該第一又は第二のハードディスクドライブを含む R A I D グループの前記第一又は第二のハードディスクドライブの中から、前記稼働時間が長い前記第一又は第二のハードディスクドライブを節電モードとすること

を特徴とするディスクアレイ装置。

【請求項 13】

請求項 8 に記載のディスクアレイ装置であって、

前記第一又は第二のハードディスクドライブ毎のヘッドがロードされた時刻をロード時刻として前記メモリに記憶するロード時刻記憶部と、

前記ロード時刻と前記計時機構より取得される時刻との差が所定の時間を超えている前記ハードディスクドライブのヘッドをアンロードするアンロード実行部と

を有することを特徴とするディスクアレイ装置。

【請求項 14】

請求項 8 に記載のディスクアレイ装置であって、

前記第一又は第二のハードディスクドライブ毎の累積稼働時間を前記メモリに記憶する累積稼働時間記憶部と、

前記第一又は第二のハードディスクドライブ毎に発生したエラーの回数を前記メモリに記憶するエラー回数記憶部と、

前記エラー回数が前記累積稼働時間に応じた所定のエラー回数を超えている前記第一又は第二のハードディスクドライブに記憶されているデータの複製を予備の前記第一又は第二のハードディスクドライブであるスペアドライブに記憶するダイナミックスペア実行部と

を有することを特徴とするディスクアレイ装置。

【請求項 15】

請求項 8 に記載のディスクアレイ装置であって、

前記情報処理装置から前記節電モードとなっている前記第一のハードディスクドライブを含んで形成される前記論理ボリュームに対する前記読み出し要求を受信すると、前記論理ボリュームを構成する前記第一のハードディスクドライブのうちの節電モードとなっていない前記第一のハードディスクドライブから前記 R A I D グループの冗長性を用いて前記読み出し要求に応答し、前記情報処理装置から前記節電モードとなっている前記第二のハードディスクドライブを含んで形成される前記論理ボリュームに対する前記読み出し要求を受信すると、節電モードとなっている前記第二のハードディスクドライブの節電モードを解除して前記読み出し要求に応答する読み出し要求応答部を有すること

を特徴とするディスクアレイ装置。

【請求項 16】

請求項 8 に記載のディスクアレイ装置であって、

前記節電モード実行部が前記第一のハードディスクドライブを節電モードとする前に、当該第一のハードディスクドライブに記憶されているデータの複製を予備の前記第一又は第二のハードディスクドライブであるスペアドライブに記憶するスペアドライブ記憶部と

、
節電モードとなっている前記第一のハードディスクドライブを含んで形成される前記論理ボリュームに対する前記書き込み要求を受信すると、前記論理ボリュームを構成する前記第一のハードディスクドライブのうちの節電モードとなっていない前記第一のハードディスクドライブと前記スペアドライブとを前記論理ボリュームが形成されている前記 R A I D グループとして前記書き込み要求に応答する書き込み要求応答部と

を有し、

前記節電モード解除部は、前記スペアドライブに記憶されているデータの複製を節電モードが解除された前記第一のハードディスクドライブに記憶すること

を特徴とするディスクアレイ装置。

【請求項 17】

請求項 16 に記載のディスクアレイ装置であって、

前記書き込み要求応答部が前記書き込み要求に応じて前記スベアドライブに書き込んだデータの位置情報を前記メモリに記憶する位置情報記憶部を有し、

前記節電モード解除部は、前記スベアドライブの前記位置情報で示される位置に記憶されている前記データの複製を節電モードが解除された前記第一のハードディスクドライブに記憶すること

を特徴とするディスクアレイ装置。

【請求項 18】

請求項 8 に記載のディスクアレイ装置であって、

節電モードとなっている前記第一のハードディスクドライブを含んで形成される前記論理ボリュームに対する前記書き込み要求を受信すると、当該書き込み要求に伴うデータを前記論理ボリュームを構成する前記第一のハードディスクドライブのうちの節電モードとなっていない前記第一のハードディスクドライブのみに書き込んで当該書き込み要求に応答する書き込み要求応答部を有し、

前記節電モード解除部は、前記論理ボリュームを構成する前記第一のハードディスクドライブのうちの節電モードとなっていなかった前記第一のハードディスクドライブに記憶されているデータから前記 R A I D グループの冗長性を用いて節電モードとなっていた前記第一のハードディスクドライブに記憶されるデータを生成し、生成された当該データを節電モードが解除された前記第一のハードディスクドライブに記憶すること

を特徴とするディスクアレイ装置。

【請求項 19】

請求項 18 に記載のディスクアレイ装置であって、

前記書き込み要求において節電モードとなっている前記第一のハードディスクドライブに書き込まれるデータの当該第一のハードディスクドライブにおける位置情報を前記メモリに記憶する位置情報記憶部を有し、

前記節電モード解除部は、前記論理ボリュームを構成する前記第一のハードディスクドライブのうちの節電モードとなっていなかった前記第一のハードディスクドライブに記憶されているデータから前記 R A I D グループの冗長性を用いて節電モードとなっていた前記第一のハードディスクドライブの前記位置情報で示される位置に記憶されるデータを生成し、生成された当該データを節電モードが解除された前記第一のハードディスクドライブに記憶すること

を特徴とするディスクアレイ装置。

【請求項 20】

情報処理装置と通信可能に接続され、

第一のインタフェース規格によりデータの送受信を行う複数の第一のハードディスクドライブと、

前記第一のハードディスクドライブよりも寿命の短い、第二のインタフェース規格によりデータの送受信を行う複数の第二のハードディスクドライブと、

前記情報処理装置から前記第一又は第二のハードディスクドライブに対するデータの読み出し要求と書き込み要求とを受信するホストインタフェースと、通信経路を介して前記第一又は第二のハードディスクドライブと通信可能に接続され前記第一又は第二のハードディスクドライブとの間でデータの入出力を行うディスクインタフェースと、メモリと、前記ホストインタフェースと前記ディスクインタフェースとの制御を司る CPU と、計時機構とを含んで構成されるコントローラとを有し、

複数の前記第一又は第二のハードディスクドライブにより構成される R A I D グループに論理ボリュームが形成されているディスクアレイ装置の制御方法であって、

前記情報処理装置から前記論理ボリュームに対するデータの読み出し要求又は書き込み要求を受信すると前記計時機構より取得される時刻をアクセス時刻として当該論理ボリュームが形成されている前記 R A I D グループの識別子と対応付けて前記メモリに記憶するステップと、

前記メモリに記憶されている前記アクセス時刻を参照し、前記計時機構より取得される時刻と前記アクセス時刻との差が所定の時間を超えているかどうか確認するステップと、

前記計時機構より取得される時刻と前記アクセス時刻との差が所定の時間を超えている前記 R A I D グループについて、前記 R A I D グループが前記第一のハードディスクドライブのみで構成されている場合は、前記 R A I D グループの冗長性に応じた数の前記第一のハードディスクドライブを節電モードとするステップと、

前記計時機構より取得される時刻と前記アクセス時刻との差が所定の時間を超えている前記 R A I D グループについて、前記 R A I D グループが前記第二のハードディスクドライブのみで構成されている場合は、任意の数の前記第二のハードディスクドライブを節電モードとするステップと、

前記第一又は第二のハードディスクドライブが節電モードとなると、前記計時機構より取得される時刻を節電開始時刻として前記 R A I D グループの識別子と対応付けて前記メモリに記憶するステップと、

前記節電開始時刻を参照し、当該節電開始時刻と前記計時機構より取得される時刻との差が所定の時間を超えている場合は、節電モードとなっている前記第一又は第二のハードディスクドライブの節電モードを解除するステップと

を有することを特徴とするディスクアレイ装置の制御方法。

【書類名】 明細書**【発明の名称】** ディスクアレイ装置及びディスクアレイ装置の制御方法**【技術分野】****【0001】**

本発明は、ディスクアレイ装置及びディスクアレイ装置の制御方法に関する。

【背景技術】**【0002】**

近年、様々な分野において、環境問題への対応やコスト削減が重要視されている。ディスクアレイ装置においては、消費する電力を減らすことにより、これらの問題に対応しようとする流れがある。特許文献1においては、ディスクアレイ装置のハードディスクドライブ群(RAIDグループ)ごとに、情報処理装置からのアクセスを管理し、一定時間アクセスが無い場合は、RAIDグループを構成する全てのハードディスクドライブを節電モードとする方法が開示されている。

【0003】

また、ディスクアレイ装置においては、個々のハードディスクドライブの障害発生を防止し、ディスクアレイ装置全体の寿命を長くすることが求められている。特許文献2においては、所定の時間ごとに、ハードディスクドライブを節電モードとする(磁気ディスクの回転を停止する)動作を1台ずつ順番に実施することにより、アクセス性能を大幅に低下させることなくハードディスクドライブの障害発生を防止する方法が開示されている。

【0004】

障害防止: 特許文献2

【特許文献1】 特開2000-293314号公報

【特許文献2】 特開平10-83614号公報

【発明の開示】**【発明が解決しようとする課題】****【0005】**

特許文献1による方法においては、RAIDグループごとに全てのハードディスクドライブを節電モードとしている。そのため、情報処理装置から当該RAIDグループに構成される論理ボリュームに対するアクセス要求があった場合に、節電モードを解除してから当該アクセス要求に応答する必要がある、アクセス性能が大幅に低下してしまう。そのため、アクセス性能を大幅に低下させることなく、ハードディスクドライブにおいて消費する電力を削減することが求められている。

【0006】

また、近年、インタフェース規格が異なる複数のハードディスクドライブを組み合わせたディスクアレイ装置が用いられるようになってきている。インタフェース規格とは、例えばファイバチャネルやシリアルATAなどである。シリアルATAのハードディスクドライブは、ファイバチャネルのハードディスクドライブと比較すると信頼性が低く、また、寿命も短い。しかし、シリアルATAのハードディスクドライブはファイバチャネルのハードディスクドライブより安価であるため、基幹業務以外の業務に用いる等、用途に応じてファイバチャネルとシリアルATAとを使い分けている場合が多い。

【0007】

例えば、ディスクアレイ装置がファイバチャネルのハードディスクドライブとシリアルATAのハードディスクドライブとにより構成されているとする。この場合、シリアルATAのハードディスクドライブの方が寿命が短いため、ディスクアレイ装置全体の寿命がシリアルATAのハードディスクドライブの寿命に影響され短くなってしまう。そのため、ディスクアレイ装置が複数のインタフェース規格のハードディスクドライブにより構成されている場合において、ハードディスクドライブの寿命を延ばすことが求められている。特許文献2に開示されている方法を用いることで、ハードディスクドライブを1台ずつ節電モードとすることによりハードディスクドライブの障害発生を防止し、ハードディスクドライブの寿命を延ばすことができるが、ディスクアレイ装置が複数のインタフェース

規格のハードディスクドライブにより構成されている場合の考慮は開示されていない。

【0008】

本発明は上記課題を鑑みてなされたものであり、ディスクアレイ装置及びディスクアレイ装置の制御方法を提供することを目的とする。

【課題を解決するための手段】

【0009】

上記課題を解決するために、本発明は、情報処理装置と通信可能に接続され、複数のハードディスクドライブと、前記情報処理装置から前記ハードディスクドライブに対するデータの読み出し要求と書き込み要求とを受信するホストインタフェースと、通信経路を介して前記ハードディスクドライブと通信可能に接続され前記ハードディスクドライブとの間でデータの入出力を行うディスクインタフェースと、メモリと、前記ホストインタフェースと前記ディスクインタフェースとの制御を司るCPUと、計時機構とを含んで構成されるコントローラとを有し、複数の前記ハードディスクドライブにより構成される冗長性を有するRAIDグループに論理ボリュームが形成され、前記情報処理装置から前記論理ボリュームに対するデータの読み出し要求又は書き込み要求を受信すると前記計時機構より取得される時刻を前記メモリにアクセス時刻として当該論理ボリュームが形成されている前記RAIDグループの識別子と対応付けて記憶するアクセス時刻記憶部と、前記メモリに記憶されている前記アクセス時刻を参照し、前記計時機構より取得される時刻と前記アクセス時刻との差が所定の時間を超えている場合、前記RAIDグループの冗長性に応じた数の前記ハードディスクドライブを節電モードとする節電モード実行部とを有することを特徴とするディスクアレイ装置に関する。

【0010】

また、本発明は、情報処理装置と通信可能に接続され、第一のインタフェース規格によりデータの送受信を行う複数の第一のハードディスクドライブと、前記第一のハードディスクドライブよりも寿命の短い、第二のインタフェース規格によりデータの送受信を行う複数の第二のハードディスクドライブと、前記情報処理装置から前記第一又は第二のハードディスクドライブに対するデータの読み出し要求と書き込み要求とを受信するホストインタフェースと、通信経路を介して前記第一又は第二のハードディスクドライブと通信可能に接続され前記第一又は第二のハードディスクドライブとの間でデータの入出力を行うディスクインタフェースと、メモリと、前記ホストインタフェースと前記ディスクインタフェースとの制御を司るCPUと、計時機構とを含んで構成されるコントローラとを有し、複数の前記第一又は第二のハードディスクドライブにより構成されるRAIDグループに論理ボリュームが形成され、前記情報処理装置から前記論理ボリュームに対するデータの読み出し要求又は書き込み要求を受信すると前記計時機構より取得される時刻をアクセス時刻として当該論理ボリュームが形成されている前記RAIDグループの識別子と対応付けて前記メモリに記憶するアクセス時刻記憶部と、前記メモリに記憶されている前記アクセス時刻を参照し、前記計時機構より取得される時刻と前記アクセス時刻との差が所定の時間を超えている前記RAIDグループについて、前記RAIDグループが前記第一のハードディスクドライブのみで構成されている場合は、前記RAIDグループの冗長性に応じた数の前記第一のハードディスクドライブを節電モードとし、前記RAIDグループが前記第二のハードディスクドライブのみで構成されている場合は、任意の数の前記第二のハードディスクドライブを節電モードとする節電モード実行部と、前記節電モード実行部により前記第一又は第二のハードディスクドライブが節電モードとなると、前記計時機構より取得される時刻を節電開始時刻として前記RAIDグループの識別子と対応付けて前記メモリに記憶する節電開始時刻記憶部と、前記節電開始時刻を参照し、当該節電開始時刻と前記計時機構より取得される時刻との差が所定の時間を超えている場合は、節電モードとなっている前記第一又は第二のハードディスクドライブの節電モードを解除する節電モード解除部とを有することを特徴とするディスクアレイ装置に関する。

【発明の効果】

【0011】

ディスクアレイ装置及びディスクアレイ装置の制御方法を提供することができる。

【発明を実施するための最良の形態】

【0012】

==第一の実施形態==

図1は、本発明の第一の実施形態として説明するディスクアレイ装置のハードウェア構成を示すブロック図である。

【0013】

ディスクアレイ装置10には、SAN (Storage Area Network) 30を介して情報処理装置20が接続されており、また、LAN (Local Area Network) を介して管理端末90が接続されている。

【0014】

情報処理装置20は、例えば、パーソナルコンピュータ、ワークステーション、メインフレームコンピュータなどである。情報処理装置10ではオペレーティングシステムが動作している。オペレーティングシステム上ではアプリケーションソフトウェアが動作している。アプリケーションソフトウェアは、例えば、銀行の自動預金預け払いシステムや航空機の座席予約システムの機能を提供する。

【0015】

管理端末90は、ディスクアレイ装置20やハードディスクドライブ80を保守・管理するためのコンピュータであり、LANを介してディスクアレイ装置10に接続されている。LANは、例えばTCP/IP等のプロトコルに従って通信が行われ、ディスクアレイ装置10と管理端末90とを接続するネットワークである。なお、管理端末90は、必ずしもLANのような通信手段で行われる必要はなく、SCSI (Small Computer Systems Interface) 規格の通信線やバスライン、ピアツーピア (Peer to Peer) 等で接続されていてもよい。また、管理端末90はディスクアレイ装置に内蔵される形態とすることもできる。

【0016】

図1に示すように、ディスクアレイ装置10には、SAN30を介して情報処理装置20が接続されている。情報処理装置20は、例えば、パーソナルコンピュータ、ワークステーション、メインフレームコンピュータなどである。

【0017】

ディスクアレイ装置10は、基本筐体11と1つ又は複数の増設筐体12を備えている。本実施の形態においては、基本筐体11は、コントローラ13、ハードディスクドライブ80などを備えている。コントローラ13は、ホストインタフェース40、ディスクインタフェース502、CPU14、メモリ15、計時機構16、キャッシュメモリ60、及びデータコントローラ17などを備えている。また、増設筐体12は、ハードディスクドライブ80などを備えている。基本筐体及び増設筐体のハードディスクドライブ80は、通信経路81により、ディスクインタフェース50と通信可能に接続されている。なお、ディスクインタフェース50とハードディスクドライブ80との接続形態の詳細については後述する。

【0018】

ホストインタフェース40は情報処理装置20との間で通信を行うインタフェースである。ホストインタフェース40は、ファイバチャネルプロトコルに従ってブロックアクセス要求を受け付ける機能を有する。

ディスクインタフェース50は、CPU14からの指示により、ハードディスクドライブ80との間でデータのやりとりを行うインタフェースである。ディスクインタフェース50は、ハードディスクドライブ80を制御するコマンドなどを規定するプロトコルに従ってハードディスクドライブ80に対するデータ入出力要求を送信する機能を備える。

【0019】

CPU14は、ディスクアレイ装置10の全体の制御を司るもので、メモリ15に格納されたマイクロプログラムを実行することにより、ホストインタフェース40、ディスク

インタフェース 50、及びデータコントローラ 17等の制御を行う。また、メモリ 15には、マイクロプログラム以外にも、後述する RAID グループ管理テーブル 301等が記憶されている。

キャッシュメモリ 60は、ホストインタフェース 40とディスクインタフェース 50との間で授受されるデータを一時的に記憶するために用いられる。

計時機構 16は、時間を計測することができる回路である。また、計時機構 16は回路等のハードウェアではなく、メモリ 15に格納されるマイクロプログラムにより実現されるものとしてもよい。

データコントローラ 17は、CPU 15の制御によりホストインタフェース 40とキャッシュメモリ 60との間、あるいはキャッシュメモリ 60とディスクインタフェース 50との間でのデータの転送を行うものである。

【0020】

コントローラ 13は、ハードディスクドライブ 80をいわゆる RAID (Redundant Array of Inexpensive Disks) 方式に規定される RAID レベル (例えば、0, 1, 5) で制御する機能を備えている。RAID 方式においては、複数のハードディスクドライブ 80が1つのグループ (以後、RAID グループと称する) として管理されている。RAID グループ 82上には、情報処理装置 20からのアクセス単位である論理ボリューム 83が形成されており、各論理ボリューム 83には LUN (Logical Unit Number) と呼ばれる識別子が付与されている。

【0021】

また、ディスクアレイ装置 10は、1つ又は複数のハードディスクドライブ 80を、RAID グループ 82を構成するハードディスクドライブ 80に障害が発生した場合の代替装置であるスペアドライブ 85として備えていることもある。

【0022】

なお、ディスクアレイ装置は、以上に説明した構成のもの以外にも、例えば、NFS (Network File System) などのプロトコルにより情報処理装置 300からファイル名指定によるデータ入出力要求を受け付けるように構成された NAS (Network Attached Storage) として機能するものなどであってもよい。

【0023】

ハードディスクドライブ 80は、例えば、FC (Fibre Channel) やシリアル ATA (Serial ATA Attachment) などのインタフェース規格により通信可能なハードディスク装置である。ハードディスクドライブ 80は、複数枚の磁気ディスクを備え、磁気ディスク上にヘッドがアクセスすることによりデータの読み書きを行う。

【0024】

ヘッドの動作方式としては、CSS (Contact Start Stop) 方式やロード／アンロード (Load/Unload) 方式などがある。CSS 方式では、磁気ディスクが回転していない時は、ヘッドは磁気ディスクの最内周部に着地しており、磁気ディスクが回転すると、回転により生じる風圧により磁気ディスクの表面にヘッドが浮上する。ロード／アンロード方式では、磁気ディスクが回転していない時は、ヘッドが磁気ディスクの外側に待避 (アンロード) され、磁気ディスクが回転すると、ヘッドが磁気ディスク上に移動 (ロード) される。

【0025】

また、ハードディスクドライブ 80は、電力消費量の異なる複数の動作モードを備えている。これら複数の動作モードは、ハードディスクドライブ 80のスピンデルの回転を制御すること等により実現されている。例えば、ハードディスクドライブ 80のインタフェース規格がシリアル ATA の場合であれば、アクティブ、アイドル、スタンバイ、及びスリープの4つの動作モードが備えられている。これら4つのモードでは、前述の順に消費電力が大きく、データ入出力要求などのアクセス要求に対する応答が速い。したがって、ハードディスクドライブ 80に対してアクセス要求が無い場合であれば、スタンバイやスリープなどの節電モードとすることにより、ハードディスクドライブ 80におえる消費電

力を減らすことができる。

【0026】

なお、ハードディスクドライブ80はディスクアレイ装置10と一体型とすることもできるし、別体とすることもできる。ハードディスクドライブ80により提供される記憶領域は、前述した論理ボリューム83を単位として管理されている。ハードディスクドライブ80へのデータの書き込みや読み出しは、論理ボリューム83に付与される識別子であるLUN (Logical Unit Number) を指定して行なうことができる。

【0027】

通信経路81は、例えばFC-A L (Fibre Channel Arbitrated Loop) などであり、ディスクインタフェース50とハードディスクドライブ80とを通信可能に接続する。なお、ディスクインタフェース50とハードディスクドライブ80とをFC-A Lハブを介して通信可能に接続することとしてもよいし、ファイバチャネルケーブルなどを直接接続することとしてもよい。

【0028】

図2は、第一の実施の形態に係るコントローラ13が備える機能を示すブロック図である。コントローラ13は、アクセス時刻記憶部201、節電モード実行部202、通信経路選択部203、読み出し要求応答部204、節電モード解除部205、スเปードライブ記憶部206、書き込み要求応答部207、位置情報記憶部208を備えている。各部201～208は、コントローラ13のCPU14がメモリ15に格納されているプログラムを実行することにより実現される。

【0029】

図3は、メモリ15に記憶されているRAIDグループ管理テーブル301及び節電待ち時間を示す図である。

【0030】

RAIDグループ管理テーブル301には、「設定RAID」、「管理LUN」、「アクセス時刻」、「実装位置」、「通信経路番号」、及び「節電開始時刻」の欄がある。「設定RAID」の欄には、RAIDグループを表す番号が設定されている。

「管理LUN」の欄には、RAIDグループ内に設定されている全ての論理ボリューム83のLUNが設定されている。「アクセス時刻」の欄には、RAIDグループ82内のいずれかの論理ボリューム83に対して情報処理装置20からアクセスがあった最後の時刻が記憶されている。ホストインタフェース40が情報処理装置20からアクセス要求を受信すると、アクセス時刻記憶部201は計時機構16から現在時刻を取得して「アクセス時刻」に設定する。「実装位置」の欄には、RAIDグループ82において節電モードとなっている全てのハードディスクドライブ80の実装されている位置が設定されている。「通信経路番号」の欄には、RAIDグループ82において節電モードとなっている全てのハードディスクドライブ80が接続されている通信経路81を示す番号が設定されている。「節電開始時刻」の欄には、RAIDグループ82においてハードディスクドライブ80が節電モードとなった時刻が設定されている。節電モード実行部202は、後述する節電処理においてハードディスクドライブ80を節電モードとする際に、「実装位置」、「通信経路番号」、及び「節電開始時刻」の節電情報の設定を行う。

【0031】

節電待ち時間302には、ハードディスクドライブ80が節電モードとなるまでの時間が設定されている。つまり、情報処理装置20から、RAIDグループ82内のいずれかの論理ボリューム80へのアクセスが、節電待ち時間302に設定されている時間無い場合は、当該RAIDグループ82において、後述する手順に従ってハードディスクドライブ80が節電モードとされる。なお、節電待ち時間302は管理端末90より設定される。

【0032】

==節電処理の流れ==

図4は、第一の実施の形態に係る節電処理を示すフローチャートである。なお、節電処

理はRAIDグループ82単位に実行される。

【0033】

節電モード実行部202は、対象としているRAIDグループ82について、RAIDグループ管理テーブル301に登録されているアクセス時刻と現在時刻との差が節電待ち時間を超過していないか監視している(S401)。節電待ち時間を超過している場合、節電モード実行部202は、当該RAIDグループ82が冗長性のあるRAID構成であるかを確認する(S402)。

【0034】

冗長性のあるRAID構成とは、RAIDグループ82を構成する複数のハードディスクドライブ80において、1つ又は複数のハードディスクドライブ80に障害が発生した場合においてもデータの読み出しを正常に行える構成となっていることである。例えば、RAID5の場合であれば、RAIDグループ82を構成する複数のハードディスクドライブ80のうちの、1つのハードディスクドライブ82に障害が発生した場合であれば、障害が発生していない残りのハードディスクドライブ82を用いて障害が発生したハードディスクドライブ82に記憶されているデータを復元することができる。ただし、RAID0においては、RAIDグループ82を構成する複数のハードディスクドライブ80のうちの1つのハードディスクドライブ80で障害が発生すると、正しくデータを読み出すことができない。そのため、RAID0においては、冗長性が無いということになる。

【0035】

節電待ち時間を超過したRAIDグループ82のRAID構成に冗長性が無い場合、節電モード実行部202は、ハードディスクドライブ80の動作モードを変更することなく処理を終了する。

【0036】

節電待ち時間を超過したRAIDグループ82のRAID構成に冗長性がある場合、節電モード実行部202は、この冗長性に応じた数のハードディスクドライブ80を節電モードとする処理を開始する。なお、冗長性に応じた数とは、この数のハードディスクドライブ80を用いなくても、残りのハードディスクドライブ80を用いてデータの読み出しを正しく行うことができる最大の数である。節電モードとすることが可能なハードディスクドライブ80の候補が複数ある場合は、通信経路選択部203は、どの通信経路81に接続されているハードディスクドライブ80を節電モードとするかを選択する(S403)。この際に、通信経路選択部203は、RAIDグループ管理テーブル301に記憶されている節電モードとなっているハードディスクドライブ80の通信経路番号を参照し、各通信経路81に接続されている節電モードとなっているハードディスクドライブ80の数が最も少ない通信経路81を選択する。節電モード実行部202は、選択された通信経路81に接続されている冗長性に応じた数のハードディスクドライブ80の動作モードを節電モードに変更する(S404)。節電モード実行部202は、節電モードとしたハードディスクドライブ80の実装位置と通信経路番号、及び節電モードとした時刻をRAIDグループ管理テーブルに登録する(S405)。

【0037】

なお、ハードディスクドライブ80を節電モードにすることは、ハードディスクドライブ80のスピンドルの回転を停止したり、動作モードをアクティブからアイドル、スタンバイ、スリープ等の状態に移すること等により実現される。

【0038】

==READ処理の流れ==

図5は、情報処理装置20からデータの読み出し要求を受信した際の、処理の流れを示すフローチャートである。

【0039】

ホストインタフェース40が情報処理装置20からデータの読み出し要求を受信すると(S501)、読み出し要求応答部204は、読み出し要求の対象となっている論理ボリューム83の属するRAIDグループ82が節電モードとなっているかどうかを確認する

(S502)。

節電モードとなっていない場合、読み出し要求応答部204は、要求のあったデータをハードディスクドライブ80から読み出し(S503)、読み出したデータを情報処理装置20に送信する。アクセス時刻記憶部201は、RAIDグループ管理テーブル301のアクセス時刻に計時機構16より取得される現在時刻を登録し(S504)、処理を終了する。

【0040】

次に、読み出し要求の対象となっている論理ボリューム83の属するRAIDグループ82が節電モードとなっている場合の処理について説明する。読み出し要求応答部204は、RAIDの冗長性を利用して節電モードとなっていないハードディスクドライブ80からデータを読み出す(S505)。読み出し要求応答部204は、読み出したデータを情報処理装置20に送信する。そして、節電モード解除部205は、節電モードとなっているハードディスクドライブ80の節電モードを解除する(S506)。アクセス時刻記憶部201は、RAIDグループ管理テーブル301のアクセス時刻に計時機構16より取得される現在時刻を登録する。節電モード解除部205は、節電情報である実装位置、通信経路番号、及び節電開始時刻を削除する(S507)。そして、節電モード解除部205は、当該読み出し要求のあったRAIDグループ82について、節電処理を開始する(S508)。

【0041】

なお、ハードディスクドライブ80の節電モードを解除することは、停止していたスピンドルの回転を再開することや、動作モードをアクティブの状態に移転すること等により実現される。

【0042】

==WRITE処理の流れ==

図6は、情報処理装置20からデータの書き込み要求を受信した際の、処理の流れを示すフローチャートである。

【0043】

ホストインタフェース40が情報処理装置20からデータの書き込み要求を受信すると(S601)、書き込み要求応答部207は、書き込み要求の対象となっている論理ボリューム83の属するRAIDグループ82が節電モードとなっているかどうかを確認する(S602)。

節電モードとなっていない場合、書き込み要求応答部207は、論理ボリューム83を構成するハードディスクドライブ80にデータを書き込むWRITE処理を実行し(S602)、情報処理装置20に書き込み処理の完了を通知する。アクセス時刻記憶部201は、RAIDグループ管理テーブル301のアクセス時刻に計時機構16より取得される現在時刻を登録し(S604)、処理を終了する。

【0044】

次に、書き込み要求の対象となっている論理ボリューム83の属するRAIDグループ82が節電モードとなっている場合の処理について説明する。ここで、図4の節電処理において、節電モード実行部202がハードディスクドライブ80を節電モードとする(S404)前に、スペアドライブ記憶部206が当該ハードディスクドライブ80に記憶されているデータの複製をスペアドライブ85に記憶しているものとする。

【0045】

書き込み要求応答部207は、論理ボリューム83を構成する節電モードとなっていないハードディスクドライブ80と当該スペアドライブ85とにデータを書き込むWRITE処理を実行し(S605)、情報処理装置20に書き込み処理の完了を通知する。

【0046】

節電モード解除部205は、節電モードとなっているハードディスクドライブ80の節電モードを解除する(S606)。節電モード解除部205は、スペアドライブ85に記憶されているデータの複製を節電モードが解除されたハードディスクドライブ80に復元

する(S607)。なお、書き込み要求応答部207がスペアドライブ85に書き込んだデータの位置情報を、位置情報記憶部208がキャッシュメモリ60に記憶しておくようにすることもできる。この場合、節電モード解除部205は、スペアドライブ85からハードディスクドライブ80にデータを復元する(S607)際に、キャッシュメモリ60に記憶されている位置情報で示されるデータのみをハードディスクドライブ80に復元することが可能となる。アクセス時刻記憶部201は、RAIDグループ管理テーブル301のアクセス時刻に計時機構16より取得される現在時刻を登録する。節電モード解除部205は、節電情報を削除する(S608)。そして、節電モード解除部205は、当該書き込み要求のあったRAIDグループ82について、節電処理を開始する(S609)。

【0047】

図6では、節電モードとなっているハードディスクドライブ80に対する書き込み要求があった場合にスペアドライブ85を用いる方法を説明したが、スペアドライブ85を用いない方法もある。その方法を図7のフローチャートにて説明する。

【0048】

図7のS701~S704の処理は図6のS601~S604の処理と同じである。書き込み要求の対象となっている論理ボリューム83の属するRAIDグループ82が節電モードとなっている場合の処理について説明する。

書き込み要求応答部207は、論理ボリューム83を構成する節電モードとなっていないハードディスクドライブ80にデータを書き込むWRITE処理を実行し(S705)、情報処理装置20に書き込み処理の完了を通知する。

【0049】

節電モード解除部205は、節電モードとなっているハードディスクドライブ80の節電モードを解除する(S706)。節電モード解除部205は、節電モードとなっていなかったハードディスクドライブ80に記憶されているデータから節電モードとなっていたハードディスクドライブ80に記憶されるデータを生成し、生成されたデータを節電モードが解除されたハードディスクドライブ80に復元する(S707)。

【0050】

なお、ハードディスクドライブ80が節電モード中に書き込まれる予定であったデータの位置情報を、位置情報記憶部208がキャッシュメモリ60に記憶しておくようにすることもできる。この場合、節電モード解除部205は、節電モードとなっていなかったハードディスクドライブ80からデータを復元する(S707)際に、キャッシュメモリ60に記憶されている位置情報で示されるデータのみをハードディスクドライブ80に復元することが可能となる。アクセス時刻記憶部201は、RAIDグループ管理テーブル301のアクセス時刻に計時機構16より取得される現在時刻を登録する。節電モード解除部205は、節電情報を削除する(S708)。そして、節電モード解除部205は、当該書き込み要求のあったRAIDグループ82について、節電処理を開始する(S709)。

以上、節電処理、READ処理、及びWRITE処理の説明を行った。

【0051】

前述の節電処理においては、RAIDグループ82に形成されている論理ボリューム83に対するデータの読み書き要求などのアクセス要求が所定の時間無い場合に、RAIDグループ82の冗長性に応じた数のハードディスクドライブ80を節電モードとしている。これにより、情報処理装置20からハードディスクドライブ80へのアクセス性能を大幅に損うことなく節電の効果を得ることができる。

【0052】

また、前述のREAD処理においては、節電モードとなっていないハードディスクドライブ80を用いてデータの読み出しを行い、その後で、節電モードとなっているハードディスクドライブ80の節電モードを解除している。このように、アクセス要求への応答に続いてハードディスクドライブ80の節電モードを解除することにより、その後に続く読

み出し要求又は書き込み要求に伴う処理を迅速に行うことができる。

【0053】

また、図6に示したWRITE処理においては、節電モードとなっているハードディスクドライブ80に書き込まれるデータをスペアドライブ85に書き込むことにより、情報処理装置20に対して書き込みが完了した旨の応答を行っている。その後で、節電モードとなっているハードディスクドライブ80の節電モードを解除し、スペアドライブ85に記憶されているデータを当該ハードディスクドライブ80に書き込んでいる。これにより、情報処理装置20は節電モードとなっているハードディスクドライブ80の節電モードが解除されるのを待たずにデータの書き込み処理を完了させることができるため、情報処理装置20からハードディスクドライブ80へのアクセス性能の低下を抑制することができる。なお、スペアドライブ85に書き込まれたデータの位置情報を記憶しておくことにより、ハードディスクドライブ80が節電モードとなっていた間にスペアドライブ85に書き込まれたデータのみをハードディスクドライブ80に復元することが可能である。これにより、ハードディスクドライブ80の節電モードを解除した際の、ハードディスクドライブ80の復元時間を短縮することができる。

【0054】

また、図7に示したWRITE処理においては、RAIDグループ82のうちの節電モードとなっていないハードディスクドライブ80のみにデータを書き込むことにより、情報処理装置20に対して書き込みが完了した旨の通知を行っている。その後で、節電モードとなっているハードディスクドライブ80の節電モードを解除し、RAIDの冗長性を利用して節電モードが解除されたハードディスクドライブ80にデータを復元している。これにより、図6のWRITE処理と同様に、情報処理装置20は節電モードとなっているハードディスクドライブ80の節電モードが解除されるのを待たずにデータの書き込み処理を完了させることができる。そのため、情報処理装置20からハードディスクドライブ80へのアクセス性能の低下を抑制することができる。なお、ハードディスクドライブ80が節電モード中に、当該ハードディスクドライブ80に書き込まれる予定であったデータの位置情報を記憶しておくこともできる。そのため、ハードディスクドライブ80が節電モードとなっていた間にRAIDグループ82の他のハードディスクドライブ80に書き込まれたデータのみから節電モードとなっていたハードディスクドライブ80のデータを復元することができる。これにより、ハードディスクドライブ80の節電モードを解除した際の、ハードディスクドライブ80の復元時間を短縮することができる。

【0055】

また、節電モードとするハードディスクドライブ80の選択の際には、通信経路選択部203が通信経路81に接続されている節電モードとなっているハードディスクドライブ80の数が最も少ない通信経路81を選択している。そのため、各通信経路81の節電モードとなっているハードディスクドライブの数が均等化される。これにより、各通信経路81においてデータのトラフィックに極端な偏りが生じて通信経路81ごとのアクセス性能に著しい差が発生することを防ぐことが可能となる。

【0056】

==第二の実施形態==

次に、本発明の第二の実施形態について説明する。第二の実施形態におけるディスクアレイ装置を含むハードウェア構成は、第一の実施形態において説明した図1により示されるブロック図と同様である。第二の実施形態においては、ディスクアレイ装置10は第一のハードディスクドライブであるインタフェース規格がファイバチャネルのハードディスクドライブ80と、第二のハードディスクドライブであるインタフェース規格がシリアルATAのハードディスクドライブ80とを有している。なお、第二のハードディスクドライブの寿命が第一のハードディスクドライブの寿命よりも短いものであれば、インタフェース規格は前述のものに限られない。

【0057】

まず、ディスクインタフェース50とハードディスクドライブ80との間の接続形態に

ついて説明する。

【0058】

図8は、ハードディスクドライブ80のインタフェース規格がファイバチャネルの場合の接続形態を示す図である。通信経路81が、FC-ALである場合、複数のPBC (Port Bypass Circuit) 801が設けられている。PBC801は、ディスクインタフェース50とインタフェース規格がファイバチャネルのハードディスクドライブ80を接続する機能を提供する。また、PBC801は、障害が発生しているハードディスクドライブ80をFC-ALから切り離し、ディスクインタフェース50と他のハードディスクドライブ80との間を通信可能とする機能を備えている。

【0059】

図9は、ハードディスクドライブ80のインタフェース規格がシリアルATA規格等のファイバチャネル以外である場合の接続形態の一例を示す図である。インタフェース規格のハードディスクドライブ80をそのままFC-ALに接続することができないため、コンバータ901を用いる。コンバータ901は、ファイバチャネルとシリアルATAのインタフェース規格における、データや信号を変換する回路である。図9の例においては、コンバータ901は、基本筐体11又は増設筐体12における全てのハードディスクドライブ80に対してファイバチャネルとシリアルATAとの変換を行うよう配設されている。そのため、当該コンバータ901が配設されている筐体においては全てのハードディスクドライブ80がシリアルATAのインタフェース規格であることとなる。

【0060】

図10は、増設筐体12にコンバータ901を図9のように配設する場合における、RAIDグループ82の構成例を示す図である。基本筐体11には、ファイバチャネルのハードディスクドライブ80により構成されるRAIDグループ1001、1002が形成されている。増設筐体12には、コンバータ901が配設されているため、シリアルATAのハードディスクドライブ80により構成されるRAIDグループ1003、1004が形成されている。

【0061】

図11は、シリアルATA規格等のファイバチャネル規格以外のハードディスクドライブ80を接続する、もう一つの例を示す図である。FC-ALには、図8と同様に複数のPBC801が接続されている。シリアルATAのハードディスクドライブ1101は、コンバータ901を備え、コンバータ901がPBC801と接続されている。これにより、シリアルATAのハードディスクドライブ1101は、ディスクインタフェース50との間でデータや信号をやりとりすることができる。また、この接続形態においては、ファイバチャネル規格のハードディスクドライブ1102を接続することも可能となる。

【0062】

図12は、コンバータ901を図11のように配設する場合における、RAIDグループ82の構成例を示す図である。基本筐体11においては、RAIDグループ1201、1102が形成されており、RAIDグループ1201は全てファイバチャネルのハードディスクドライブ80、RAIDグループ1202は全てシリアルATAのハードディスクドライブ80となっている。また、増設筐体12においては、RAIDグループ1203、1204が形成されており、いずれもファイバチャネルのハードディスクドライブ80とシリアルATAのハードディスクドライブ80との混成により構成されている。このように、コンバータ901を図11のように配設する場合、一つの筐体内、また、一つのRAIDグループ82内において、ファイバチャネルのハードディスクドライブ80とシリアルATAのハードディスクドライブ80とを混在させることが可能となる。

【0063】

図13は、第二の実施形態に係るコントローラ13が備える機能を示すブロック図である。コントローラ13は、アクセス時刻記憶部1301、節電モード実行部1302、節電開始時刻記憶部1303、節電モード解除部1304、連続稼働時間記憶部1305、一括スベア実行部1306、累積稼働時間記憶部1307、ロード自国機億部1308、

アンロード実行部1309、エラー回数記憶部1310、ダイナミックスペア実行部1311、読み出し要求応答部1312、スペアドライブ記憶部1313、書き込み要求応答部1314、位置情報記憶部1315、上限チェック部1316を備えている。各部1301～1316は、コントローラ13のCPU14がメモリ15に格納されているプログラムを実行することにより実現される。

【0064】

連続稼働時間記憶部1305は、RAIDグループ82ごとの連続稼働時間をRAIDグループ管理テーブル1401に記憶する。RAIDグループ82ごとの連続稼働時間とは、RAIDグループ82に属するハードディスクドライブ80が節電モードとならずに連続して稼働している時間である。

累積稼働時間記憶部1307は、ハードディスクドライブ80ごとの累積稼働時間をハードディスクドライブ管理テーブル1402に記憶する。ハードディスクドライブ80ごとの累積稼働時間とは、ハードディスクドライブ80が節電モードとならずに稼働している時間の累積である。

ロード時刻記憶部1308は、ハードディスクドライブ80のヘッドの動作方式がロード／アンロード方式である場合に、ヘッドを磁気ディスク上にロードする際に計時機構16より現在時刻を取得してハードディスクドライブ管理テーブル1402に記憶する。

エラー回数記憶部1310は、ハードディスクドライブ80において発生したエラーの回数を、ハードディスクドライブ80ごとにハードディスクドライブ管理テーブル1402に記憶する。

【0065】

図14は、メモリ15に記憶されているRAIDグループ管理テーブル1401、ハードディスクドライブ管理テーブル1402、節電管理テーブル1403、スペア管理テーブル1404、アンロード待ち時間1405、上限管理テーブル1406、及びエラー管理テーブル1407を示す図である。

【0066】

RAIDグループ管理テーブル1401は、第一の実施形態と同様に「設定RAID」、「管理LUN」、「節電開始時刻」、及び「アクセス時刻」の欄を備えている。これらの欄に設定される内容は、第一の実施形態と同じである。

【0067】

RAIDグループ管理テーブル1401は、さらに、「ドライブタイプ」、「利用形態」、及び「連続稼働時間」の欄を備えている。「ドライブタイプ」の欄には、RAIDグループ82を構成するハードディスクドライブ80のインタフェース規格の情報が設定されている。本実施の形態においては、RAIDグループ82に属する全てのハードディスクドライブ80のインタフェース規格がファイバチャネルの場合は「FC」、RAIDグループ82に属する全てのハードディスクドライブ80のインタフェース規格がシリアルATAの場合は「SATA」と設定されている。また、図12のように、RAIDグループ82に属するハードディスクドライブ80のインタフェースが、ファイバチャネルとシリアルATAとの混在である場合は、「混在」と設定されている。「利用形態」の欄にはRAIDグループ82ごとの利用形態が設定されている。

利用形態とは、RAIDグループ82が基幹業務等のアクセス性能が非常に重要となる処理において利用されているか、またはバックアップ処理等のアクセス性能の低下が重大な問題とならないような処理において利用されているかを示すものである。本実施の形態においては、基幹業務等における利用形態を「オンライン」、バックアップ処理等における利用形態を「ニアライン」と表す。

【0068】

ハードディスクドライブ管理テーブル1402は、ハードディスクドライブ80ごとの情報を管理するものであり、「設定RAID」、「ドライブ番号」、「ロード時刻」、「累積稼働時間」、「アンロード回数」、及び「エラー回数」の欄を備えている。

「設定RAID」の欄には、ハードディスクドライブ80が属するRAIDグループ8

2が設定されている。「ドライブ番号」の欄には、そのハードディスクドライブ80を示す番号が設定されている。「ロード時刻」の欄には、ロード時刻記憶部1308により登録される、磁気ディスク上にヘッドがロードされた時刻が設定されている。「累積稼働時間」の欄には、累積稼働時間記憶部1307により登録されるハードディスクドライブ80ごとの累積稼働時間が設定されている。「アンロード回数」の欄には、ハードディスクドライブ80において、ヘッドがアンロードされた回数が記憶されている。「エラー回数」の欄には、エラー回数記憶部1310により登録される、ハードディスクドライブ80における書き込みエラー等のエラー回数が記憶されている。

【0069】

節電管理テーブル1403は、「節電待ち時間」と「節電時間」の欄を備えている。「節電待ち時間」の欄には、第一の実施形態と同様に、ハードディスクドライブ80が節電モードとなるまでの時間が設定されている。「節電時間」の欄には、節電モードとなっているハードディスクドライブ80の節電モードを解除するまでの時間が設定されている。

【0070】

スベア管理テーブル1404は、「スベア待ち時間」と「スベア時間」の欄を備えている。「スベア待ち時間」の欄には、RAIDグループ82単位でハードディスクドライブ80に記憶されているデータをスベアドライブ85にコピーし、当該RAIDグループ82の全てのハードディスクドライブ80を節電モードとするまでの待ち時間が設定されている。「スベア時間」の欄には、スベアドライブ85を用いることにより節電モードとなっているハードディスクドライブ80の節電モードを解除するまでの時間が設定されている。

【0071】

アンロード待ち時間1405には、ハードディスクドライブ80の磁気ディスク上にヘッドがロードされたロード時刻からヘッドをアンロードするまでの待ち時間が設定されている。上限管理テーブル1406は、「上限稼働時間」及び「上限アンロード回数」の欄を備えている。ハードディスクドライブ管理テーブル1402に記憶されている、ハードディスクドライブ80ごとの累積稼働時間またはアンロード回数が、上限管理テーブル1406に設定されている上限稼働時間または上限アンロード回数を超えると、当該ハードディスクドライブ80の交換要求が管理端末90に送信される。

【0072】

エラー管理テーブル1407は、「稼働時間」及び「上限エラー回数」の欄を備えている。「稼働時間」の欄には、ハードディスクドライブ80における累積稼働時間の範囲が設定されており、「上限エラー回数」の欄には、稼働時間に応じたエラー回数の上限値が設定されている。ハードディスクドライブ80のエラー回数が、エラー管理テーブル1407に設定されている稼働時間に応じた上限エラー回数を超えると、当該ハードディスクドライブ80について後述するダイナミックスベア処理が行われる。

【0073】

==節電処理の流れ==

図15は、CPU14が実行する節電処理を示すフローチャートである。なお、節電処理はRAIDグループ82単位に実行される。

【0074】

節電モード実行部1302は、対象としているRAIDグループ82について、RAIDグループ管理テーブル301に登録されているアクセス時刻と計時機構16より取得される現在時刻との差が節電管理テーブル1403に設定されている節電待ち時間を超過していないか監視している(S1501)。節電待ち時間を超過している場合、節電モード実行部1302は、当該RAIDグループ82のドライブタイプが「FC」であるか、または利用形態が「オンライン」であるかを確認する(S1502)。

【0075】

節電モード実行部1302は、ドライブタイプが「FC」または利用形態が「オンライン」である場合、当該RAIDグループ82の冗長性に応じた数のハードディスクドライブ

80を、ハードディスクドライブ管理テーブル1402を参照して累積稼働時間が長いものを優先して選択する(S1503)。節電モード実行部1302は、選択されたハードディスクドライブ80を節電モードとする(S1504)。節電開始時刻記憶部1303は、計時機構16より取得した現在時刻を節電開始時刻としてRAIDグループ管理テーブル1401に登録する(S1505)。節電モード解除部1304は、節電開始時刻と計時機構16より取得される現在時刻との差が節電管理テーブル1403に設定されている節電時間を超過していないか監視している(S1506)。節電時間を超過すると、節電モード解除部1304は、節電モードとなっているハードディスクドライブ80の節電モードを解除し(S1507)、RAIDグループ管理テーブル1401の節電開始時刻を削除する(S1508)。ドライブタイプが「FC」または利用形態が「オンライン」である場合、累積稼働時間の長いハードディスクドライブ80を選択し、節電モードとする一連の処理(S1503~S1508)は繰り返し実行される。

【0076】

ドライブタイプが「FC」以外で、利用形態が「ニアライン」の場合、節電モード実行部1302は、RAIDグループ82の冗長性によらず、任意の数のハードディスクドライブ80を節電モードとする。本実施の形態においては、節電モード実行部1302は当該RAIDグループ82に属する全てのハードディスクドライブ80を節電モードとする(S1509)。なお、節電モードとするハードディスクドライブ80は、RAIDグループに属する全てでなくてもよい。

【0077】

==READ処理の流れ==

図16は、情報処理装置20からデータの読み出し要求を受信した際の処理の流れを示すフローチャートである。

【0078】

ホストインタフェース40が情報処理装置20からデータの読み出し要求を受信すると(S1601)、読み出し要求応答部1312は、RAIDグループ管理テーブル1401を参照し、読み出し要求の対象となっている論理ボリューム83の属するRAIDグループ82が節電モードとなっているかどうかを確認する(S1602)。節電モードとなっている場合、読み出し要求応答部1312は、節電モードとなっているRAIDグループ82のドライブタイプが「FC」であるか、または利用形態が「オンライン」であるかを確認する(S1603)。

【0079】

節電モードとなっていない場合、及び節電モードとなっているがドライブタイプが「FC」であるか又は利用形態が「オンライン」である場合、読み出し要求応答部1312は、節電モードとなっていないハードディスクドライブ80からデータを読み出し(S1604)、情報処理装置20にデータを送信する。そして、読み出し要求応答部1312は、RAIDグループ管理テーブル1401のアクセス時刻に計時機構16より取得される現在時刻を登録し(S1605)、処理を終了する。

【0080】

次に、読み出し要求の対象となっている論理ボリューム83の属するRAIDグループ82が節電モードであり、かつ、ドライブタイプが「FC」でなく利用形態が「オンライン」でない場合の処理について説明する。この場合、RAIDグループ82に属する全てのハードディスクドライブ80が節電モードとなっている。そのため、読み出し要求応答部1312は、これらのハードディスクドライブ80の節電モードを解除する(S1606)。そして、読み出し要求応答部1312は節電モードが解除されたハードディスクドライブ80からデータを読み出し(S1607)、情報処理装置20にデータを送信する。アクセス時刻記憶部1312は、RAIDグループ管理テーブル1401のアクセス時刻に計時機構16より取得される現在時刻を登録する。読み出し要求応答部1312は、節電開始時刻を削除する(S1608)。そして、読み出し要求応答部1312は当該読み出し要求のあったRAIDグループについて、節電処理を開始する(S1609)。

【0081】

==WRITE 処理の流れ==

図17は、情報処理装置20からデータの書き込み要求を受信した際の処理の流れを示すフローチャートである。

【0082】

ホストインタフェース40が情報処理装置20からデータの書き込み要求を受信すると(S1701)、書き込み要求応答部1314は、書き込み要求の対象となっている論理ボリューム83の属するRAIDグループ82が節電モードとなっているかどうかを確認する(S1702)。

【0083】

節電モードとなっていない場合、書き込み要求応答部1314は、データを論理ボリューム83を構成するハードディスクドライブ80に書き込むWRITE処理を実行し(S1703)、情報処理装置20に書き込み処理の完了を通知する。アクセス時刻記憶部201は、RAIDグループ管理テーブル301のアクセス時刻に計時機構16より取得される現在時刻を登録し(S1704)、処理を終了する。

【0084】

書き込み要求の対象となっている論理ボリューム83の属するRAIDグループ82が節電モードとなっている場合の処理について説明する。書き込み要求応答部1314は、節電モードとなっているRAIDグループ82のドライブタイプが「FC」であるか、または利用形態が「オンライン」であるかを確認する(S1705)。

【0085】

まず、書き込み要求の対象となっている論理ボリューム83の属するRAIDグループ82のドライブタイプが「FC」でなく利用形態が「オンライン」でない場合の処理について説明する。書き込み要求応答部1314は、節電モードとなっているハードディスクドライブ80の節電モードを解除する(S1706)。書き込み要求応答部1314は、節電モードが解除されたハードディスクドライブ80を用いてWRITE処理を実行し(S1707)、情報処理装置20に書き込み処理の完了を通知する。アクセス時刻記憶部1301は、RAIDグループ管理テーブル301のアクセス時刻に計時機構16より取得される現在時刻を登録する。そして、書き込み要求応答部1314は、当該RAIDグループに対する節電処理を開始する(S1709)。

【0086】

次に、書き込み要求の対象となっている論理ボリューム83の属するRAIDグループ82のドライブタイプが「FC」であるか又は利用形態が「オンライン」である場合の処理について説明する。ここで、図15の節電処理において、節電モード実行部1302がRAIDグループ82の冗長性に応じた数のハードディスクドライブ80を節電モードとする(S1504)前に、スペアドライブ記憶部1313が当該ハードディスクドライブ80に記憶されているデータの複製をスペアドライブ85に記憶しているものとする。書き込み要求応答部1314は、論理ボリューム83を構成する節電モードとなっていないハードディスクドライブ80と当該スペアドライブ85とにデータを書き込むWRITE処理を実行し(S1710)、情報処理装置20に書き込み処理の完了を通知する。位置情報記憶部1315は、スペアドライブ85に書き込まれたデータの位置情報をキャッシュメモリ60に記憶する(S1711)。アクセス時刻記憶部1301は、RAIDグループ管理テーブル301のアクセス時刻に計時機構16より取得される現在時刻を登録する(S1712)。

【0087】

このように、スペアドライブ85が用いられる場合、図15の節電処理で節電モードが解除される(S1507)際に、節電モード解除部1304は、スペアディスク85に記憶されているデータをハードディスクドライブ80に復元する。つまり、節電モード解除部1304はスペアディスク85に記憶されているデータの複製を節電モードが解除されたハードディスクドライブ80に復元する。この際、節電モード解除部1304は、キャ

ッシュメモリ 60 に記憶されている位置情報で示されるデータのみを復元することとしてもよい。これにより、ハードディスクドライブ 80 へのデータの復元時間を短縮することが可能となる。

【0088】

図 17 では、スベアドライブ 85 を用いる方法を説明したが、スベアドライブ 85 を用いない方法もある。その方法を図 18 のフローチャートにて説明する。

図 18 の S1801～S1809 の処理は図 17 の S1701～S1709 の処理と同じである。

【0089】

書き込み要求の対象となっている論理ボリューム 83 の属する RAID グループ 82 が節電モードであり、ドライブタイプが「FC」であるか又は利用形態が「オンライン」である場合の処理について説明する。

書き込み要求応答部 1314 は、論理ボリューム 83 を構成する節電モードとなっていないハードディスクドライブ 80 にデータを書き込む WRITE 処理を実行し (S1810)、情報処理装置 20 に書き込み処理の完了を通知する。位置情報記憶部 1315 は、節電モードとなっているハードディスクドライブ 80 に書き込まれる予定であったデータの位置情報をキャッシュメモリ 60 に記憶する (S1811)。アクセス時刻記憶部 1301 は、RAID グループ管理テーブル 301 のアクセス時刻に計時機構 16 より取得される現在時刻を登録する (S1812)。

【0090】

このように、RAID グループ 82 の節電モードとなっていないハードディスクドライブ 80 のみにデータが書き込まれる場合、図 15 の節電処理で節電モードが解除される (S1507) 際に、節電モード解除部 1304 は、RAID の冗長性を利用して節電モードとなっていたハードディスクドライブ 80 にデータを復元する。つまり、節電モード解除部 1304 は RAID グループの節電モードとなっていなかったハードディスクドライブ 80 に記憶されているデータから節電モードとなっていたハードディスクドライブ 80 のデータを生成する。そして、節電モード解除部 1304 は、生成されたデータを節電モードが解除されたハードディスクドライブ 80 に記憶する。この際、節電モード解除部 1304 は、キャッシュメモリ 60 に記憶されている位置情報で示されるデータのみを復元することとしてもよい。これにより、ハードディスクドライブ 80 へのデータの復元時間を短縮することが可能となる。

【0091】

以上、節電処理、READ 処理、及び WRITE 処理の説明を行った。シリアル ATA のハードディスクドライブ 80 はファイバチャネルのハードディスクドライブ 80 よりも信頼性が低く、また、寿命も短い。しかし、シリアル ATA のハードディスクドライブ 80 はファイバチャネルのハードディスクドライブ 80 と比較して安価であるため、基幹業務以外における利用が急速に進んでいる。また、シリアル ATA のハードディスクドライブ 80 においては、稼働時間が寿命を大きく左右するため、稼働時間を抑制することにより耐用年数を延ばすことが可能である。そのため、前述した節電処理のように、シリアル ATA のハードディスクドライブ 80 においては、ファイバチャネルのハードディスクドライブ 80 よりも多くの数のハードディスクドライブ 80 を節電モードとすることにより、シリアル ATA のハードディスクドライブ 80 により構成されるディスクアレイ装置 10 全体の寿命を延ばすことが可能となる。なお、インタフェース規格としてファイバチャネルとシリアル ATA とを用いて説明したが、寿命が異なるものであれば、同様の効果を得ることができる。

【0092】

また、ファイバチャネルのハードディスクドライブ 80 は、基幹業務のオンライン処理など、アクセス性能が重要となる業務に用いられていることが多い。そのため、RAID グループ 82 の冗長性に応じた数のハードディスクドライブ 80 を節電モードとすることにより、アクセス性能を大幅に損なうことなく節電効果を得ることができる。また、シリ

アル A T A のハードディスクドライブ 8 0 は、基幹業務以外の業務など、アクセス性能の低下が大きな問題とならない業務に用いられることが多い。そのため、R A I D グループ 8 2 の冗長性によらず、例えば R A I D グループ 8 2 を構成する全てのハードディスクドライブ 8 0 を節電モードとすることにより、より大きな節電効果を得ることができる。

【0093】

また、ファイバチャネルのハードディスクドライブ 8 0 とシリアル A T A のハードディスクドライブ 8 0 との混成により構成される R A I D グループ 8 2 が基幹業務の「オンライン」処理など、アクセス性能が重要となる業務に用いられる場合もある。この場合は、R A I D グループ 8 2 の冗長性に応じた数のハードディスクドライブを節電モードとすることにより、アクセス性能を大幅に低下させずに節電効果を得ることができる。また、R A I D グループ 8 2 が基幹業務以外の処理、いわゆる「ニアライン」処理などのアクセス性能の低下が大きな問題とならない業務に用いられる場合もある。この場合は、R A I D グループ 8 2 の冗長性によらず、例えば R A I D グループ 8 2 を構成する全てのハードディスクドライブ 8 0 を節電モードとすることにより、より大きな節電効果を得ることができる。

【0094】

また、R A I D グループ 8 2 の冗長性に応じた数のハードディスクドライブ 8 0 を節電モードとする場合においては、R A I D グループ 8 2 に属するハードディスクドライブ 8 0 の中で、累積稼働時間が長いハードディスクドライブ 8 0 を節電モードとしている。これにより、R A I D グループ 8 2 内におけるハードディスクドライブ 8 0 の累積稼働時間のばらつきを抑え、ハードディスクドライブ 8 0 の寿命を平準化することができる。また、ハードディスクドライブ 8 0 が節電モードとなっている状態において、情報処理装置 2 0 から当該ハードディスクドライブ 8 0 が属する R A I D グループ 8 2 に対する読み出し要求又は書き込み要求が無い状態が続く場合においては、当該 R A I D グループ 8 2 内において、所定の時間ごとに節電モードとするハードディスクドライブ 8 0 を変更している。これにより、R A I D グループ 8 2 内におけるハードディスクドライブ 8 0 の累積稼働時間のばらつきを抑え、ハードディスクドライブ 8 0 の寿命を平準化することができる。

【0095】

==スベア処理の流れ==

図 19 は、スベアドライブ 8 5 を用いてハードディスクドライブ 8 0 を R A I D グループ単位で節電モードとする処理の流れを示すフローチャートである。

【0096】

一括スベア実行部 1 3 0 6 は、R A I D グループ管理テーブル 1 4 0 1 を参照し、連続稼働時間がスベア管理テーブル 1 4 0 4 に設定されているスベア待ち時間を超過している R A I D グループ 8 2 がないか監視している (S 1 9 0 1)。スベア待ち時間を超過している R A I D グループ 8 2 がある場合、一括スベア実行部 1 3 0 6 は、当該 R A I D グループ 8 2 に属する全てのハードディスクドライブ 8 0 に記憶されているデータの複製をスベアドライブ 8 5 に記憶する (S 1 9 0 2)。この状態において、情報処理装置 2 0 から当該 R A I D グループに対する読み出し要求が生じた場合、読み出し要求応答部 1 3 1 2 は、スベアドライブ 8 5 を用いて読み出し要求に対する処理を実行する。同様に、情報処理装置 2 0 から当該 R A I D グループに対する書き込み要求が生じた場合、書き込み要求応答部 1 3 1 4 は、スベアドライブ 8 5 を用いて書き込み要求に対する処理を実行する。

【0097】

一括スベア実行部 1 3 0 6 は、当該 R A I D グループ 8 2 に属する全てのハードディスクドライブ 8 0 を節電モードとする (S 1 9 0 3)。一括スベア実行部 1 3 0 6 は、計時機構 1 6 より取得される現在時刻を節電開始時刻として R A I D グループ管理テーブル 1 4 0 1 に登録する (S 1 9 0 4)。そして、一括スベア実行部 1 3 0 6 は、節電開始時刻と現在時刻との差がスベア管理テーブル 1 4 0 4 に設定されているスベア時間を超過していないか監視する (S 1 9 0 5)。スベア時間を超過すると、一括スベア実行部 1 3 0 6 は、当該 R A I D グループ 8 2 の全てのハードディスクドライブ 8 0 の節電モードを解除

する (S1906)。そして、一括スベア実行部1306は、スベアドライブ85に記憶されているデータの複製を当該RAIDグループ82のハードディスクドライブ80に記憶する。一括スベア実行部1306は、RAIDグループ管理テーブル1401の節電開始時刻を削除し (S1907)、スベア待ち時間を超過しているRAIDグループ82の監視を再開する (S1901)。

【0098】

このように、アクセス要求の有無にかかわらず、一定時間ごとにRAIDグループ82単位でスベアドライブ85にデータをコピーした上でハードディスクドライブ80を節電モードとすることにより、情報処理装置20からのアクセス性能を低下させることなく、ハードディスクドライブ80の寿命を延ばすことができる。

【0099】

==アンロード処理の流れ==

図20は、ハードディスクドライブ80のヘッドをアンロードする処理の流れを示すフローチャートである。

【0100】

アンロード実行部1309は、ハードディスクドライブ管理テーブル1402を参照し、ロード時刻と計時機構16より取得される現在時刻との差がアンロード待ち時間1405に設定されている時間を超過しているハードディスクドライブ80がないか監視している (S2001)。アンロード待ち時間を超過しているハードディスクドライブ80がある場合、アンロード実行部1309は、当該ハードディスクドライブ80のヘッドをアンロードする (S2002)。アンロード実行部1309は、当該ハードディスクドライブ80について、ハードディスクドライブ管理テーブルのアンロード回数に1加算する (S2003)。そして、アンロード実行部1309は、当該ハードディスクドライブ80のヘッドを再度ロードする (S2003)。ロード時刻記憶部1308は、ハードディスクドライブ管理テーブル1402のロード時刻に計時機構16より取得される現在時刻を登録する (S2004)。

【0101】

ハードディスクドライブ80の動作状態においては、回転する磁気ディスク上に数十nm程の隙間をあけてデータの読み書きを行うヘッドが位置している。磁気ディスクの回転中は、ヘッドからの風圧により、磁気ディスク表面にある潤滑剤に溝が生じてしまう。そのため、ロード／アンロード方式のハードディスクドライブ80の場合であれば、一定時間毎にヘッドをアンロードすることにより、磁気ディスク表面にある潤滑剤の凹凸を均し、外部からの衝撃等により磁気ディスクに障害が発生する可能性を減らすことができる。

【0102】

==上限チェック処理の流れ==

図21は、ハードディスクドライブ80の寿命を判定し、必要に応じて管理端末90にハードディスクドライブ80の交換要求を送信する処理の流れを示すフローチャートである。

【0103】

上限チェック部1316は、ハードディスクドライブ管理テーブル1402を参照し、累積稼働時間が上限管理テーブル1406に設定されている上限稼働時間を超えている、または、アンロード回数が上限管理テーブル1406に設定されている上限アンロード回数を超えているハードディスクドライブ80がないか監視している (S2001)。上限稼働時間または上限アンロード時間を超過しているハードディスクドライブ80がある場合、上限チェック部1316は、当該ハードディスクドライブ80が寿命であることを管理端末90に送信する。

【0104】

このように、累積稼働時間やヘッドのアンロード回数をもとに、ハードディスクドライブ80の寿命を判断し、管理端末90等にハードディスクドライブ80の交換要求を送信することにより、ハードディスクドライブ80の障害を未然に防ぐことができる。これに

より、ディスクアレイ装置 10 全体における可用性を高めることができる。

【0105】

==ダイナミックスペア処理の流れ==

図 22 は、ハードディスクドライブ 80 において発生しているエラーの回数をもとに、ハードディスクドライブ 80 をスペアドライブ 85 にて代替する処理の流れを示すフローチャートである。

【0106】

ダイナミックスペア実行部 1311 は、ハードディスクドライブ管理テーブル 1402 を参照し、エラー回数がエラー管理テーブルに設定されている、累積稼働時間に応じた上限エラー回数を超過しているハードディスクドライブ 80 がないか監視している (S2201)。上限エラー回数を超過しているハードディスクドライブ 80 がある場合、ダイナミックスペア実行部 1311 は、当該ハードディスクドライブ 80 に記憶されているデータの複製をスペアドライブ 85 にコピーする (S2202)。

【0107】

この状態において、情報処理装置 20 から当該ハードディスクドライブ 80 を含んで構成される論理ボリューム 83 に対する読み出し要求が生じた場合、読み出し要求応答部 1312 は、スペアドライブ 85 を用いて読み出し要求に対する処理を実行する。同様に、情報処理装置 20 から当該ハードディスクドライブ 80 を含んで構成される論理ボリューム 83 に対する書き込み要求が生じた場合、書き込み要求応答部 1314 は、スペアドライブ 85 を用いて書き込み要求に対する処理を実行する。

【0108】

これにより、累積稼働時間ごとにエラーの上限値を設定し、当該上限値を超えた場合にスペアドライブ 85 にデータをコピーすることができる。したがって、ハードディスクドライブ 80 に障害が発生する前に、スペアドライブ 85 を用いて情報処理装置 20 からのアクセス要求への応答を開始する等、ハードディスクドライブ 80 の障害によるアクセスの中断を回避するための制御を行うことが可能となる。

【0109】

以上、本実施の形態について説明したが、上記実施例は本発明の理解を容易にするためのものであり、本発明を限定して解釈するためのものではない。本発明は、その趣旨を逸脱することなく、変更、改良され得ると共に、本発明にはその等価物も含まれる。

【図面の簡単な説明】

【0110】

【図 1】第一の実施形態および第二の実施形態における、ディスクアレイ装置のハードウェア構成を示すブロック図である。

【図 2】第一の実施の形態における、コントローラが備える機能を示すブロック図である。

【図 3】第一の実施形態における、RAID グループ管理テーブルおよび節電待ち時間を示す図である。

【図 4】第一の実施形態における、節電処理を示すフローチャートである。

【図 5】第一の実施形態における、READ 処理を示すフローチャートである。

【図 6】第一の実施形態における、スペアドライブを用いた WRITE 処理を示すフローチャートである。

【図 7】第一の実施形態における、スペアドライブを用いない WRITE 処理を示すフローチャートである。

【図 8】第一の実施形態における、ファイバチャネルのハードディスクドライブの接続形態を示す図である。

【図 9】第一の実施形態における、シリアル ATA のハードディスクドライブの第一の接続形態を示す図である。

【図 10】第一の実施形態における、増設筐体にコンバータを図 9 のように配設する場合における、RAID グループの構成例を示す図である。

【図 1 1】第一の実施形態における、シリアル A T A のハードディスクドライブの第二の接続形態を示す図である。

【図 1 2】第一の実施形態における、コンバータを図 1 1 のように配設する場合における、R A I D グループの構成例を示す図である。

【図 1 3】第二の実施形態における、コントローラが備える機能を示すブロック図である。

【図 1 4】第二の実施形態における、R A I D グループ管理テーブル、ハードディスクドライブ管理テーブル、節電管理テーブル、スペア管理テーブル、アンロード待ち時間、上限管理テーブル、及びエラー管理テーブルを示す図である。

【図 1 5】第二の実施形態における、節電処理を示すフローチャートである。

【図 1 6】第二の実施形態における、R E A D 処理を示すフローチャートである。

【図 1 7】第二の実施形態における、スペアドライブを用いた W R I T E 処理を示すフローチャートである。

【図 1 8】第二の実施形態における、スペアドライブを用いない W R I T E 処理を示すフローチャートである。

【図 1 9】第二の実施形態における、スペアドライブを用いてハードディスクドライブを R A I D グループ単位で節電モードとする処理を示すフローチャートである。

【図 2 0】第二の実施形態における、ハードディスクドライブのヘッドをアンロードする処理を示すフローチャートである。

【図 2 1】第二の実施形態における、ハードディスクドライブの寿命を判定する処理を示すフローチャートである。

【図 2 2】第二の実施形態における、ハードディスクドライブにおいて発生しているエラーの回数をもとに、ハードディスクドライブをスペアドライブにて代替する処理を示すフローチャートである。

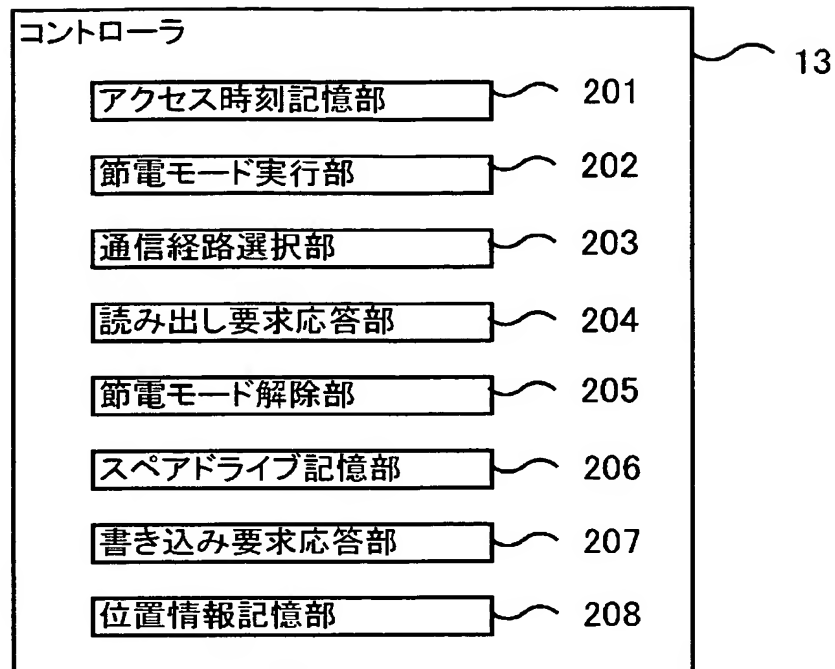
【符号の説明】

【0 1 1 1】

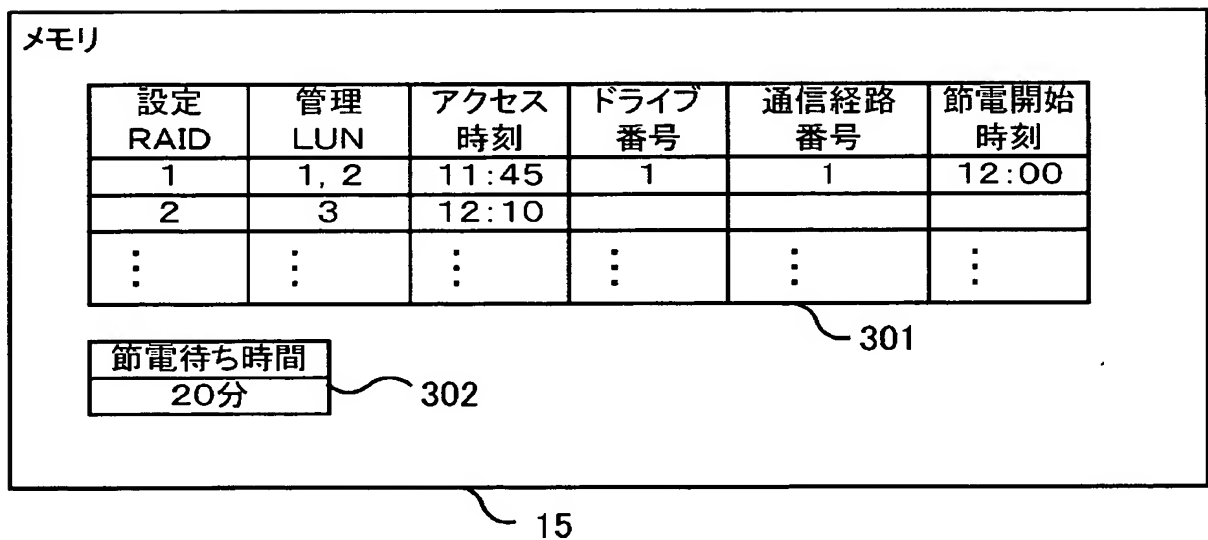
1 0	ディスクアレイ装置	1 1	基本筐体
1 2	増設筐体	1 3	コントローラ
1 4	C P U	1 5	メモリ
1 6	計時機構	1 7	データコントローラ
2 0	情報処理装置	3 0	S A N
4 0	ホストインタフェース	5 0	ディスクインタフェース
6 0	キャッシュメモリ	8 0	ハードディスクドライブ
8 1	通信経路	8 2	R A I D グループ
8 3	論理ボリューム	8 5	スペアドライブ
9 0	管理端末		
2 0 1	アクセス時刻記憶部	2 0 2	節電モード実行部
2 0 3	通信経路選択部	2 0 4	読み出し要求応答部
2 0 5	節電モード解除部	2 0 6	スペアドライブ記憶部
2 0 7	書き込み要求応答部	2 0 8	位置情報記憶部
3 0 1	R A I D グループ管理テーブル	3 0 2	節電待ち時間
8 0 1	P B C	9 0 1	コンバータ
1 3 0 1	アクセス時刻記憶部	1 3 0 2	節電モード実行部
1 3 0 3	節電開始時刻記憶部	1 3 0 4	節電モード解除部
1 3 0 5	連続稼働時間記憶部	1 3 0 6	一括スペア実行部
1 3 0 7	累積稼働時間記憶部	1 3 0 8	ロード時刻記憶部
1 3 0 9	アンロード実行部	1 3 1 0	エラー回数記憶部
1 3 1 1	ダイナミックスペア実行部	1 3 1 2	読み出し要求応答部
1 3 1 3	スペアドライブ記憶部	1 3 1 4	書き込み要求応答部
1 3 1 5	位置情報記憶部	1 3 1 6	上限チェック部

- 1 4 0 1 R A I Dグループ管理テーブル
- 1 4 0 2 ハードディスクドライブ管理テーブル
- 1 4 0 3 節電管理テーブル
- 1 4 0 4 スペア管理テーブル
- 1 4 0 5 アンロード待ち時間
- 1 4 0 6 上限管理テーブル
- 1 4 0 7 エラー管理テーブル

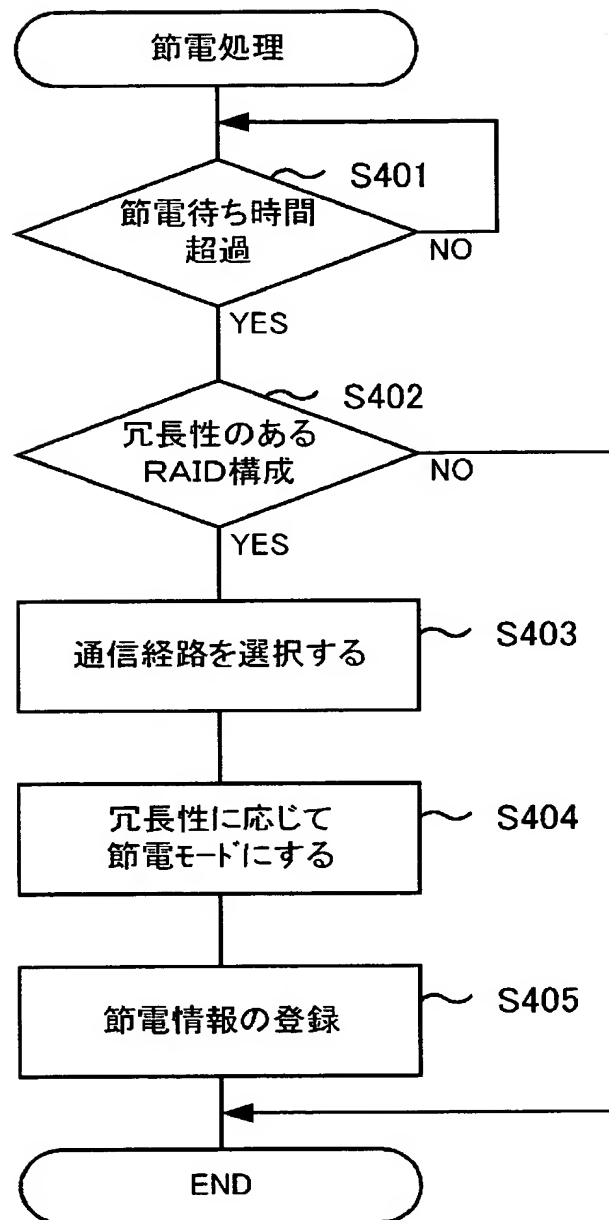
【図 2】



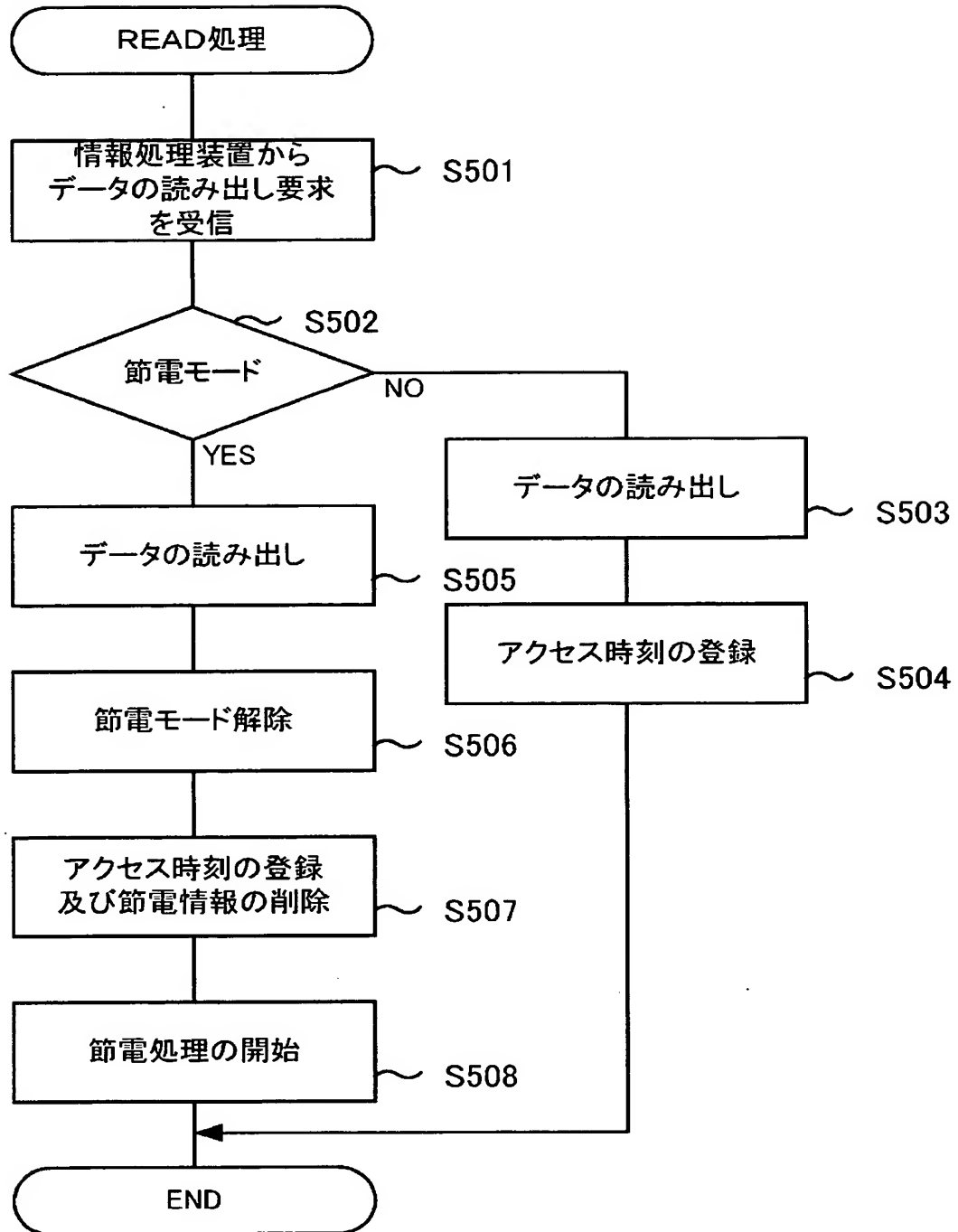
【図 3】



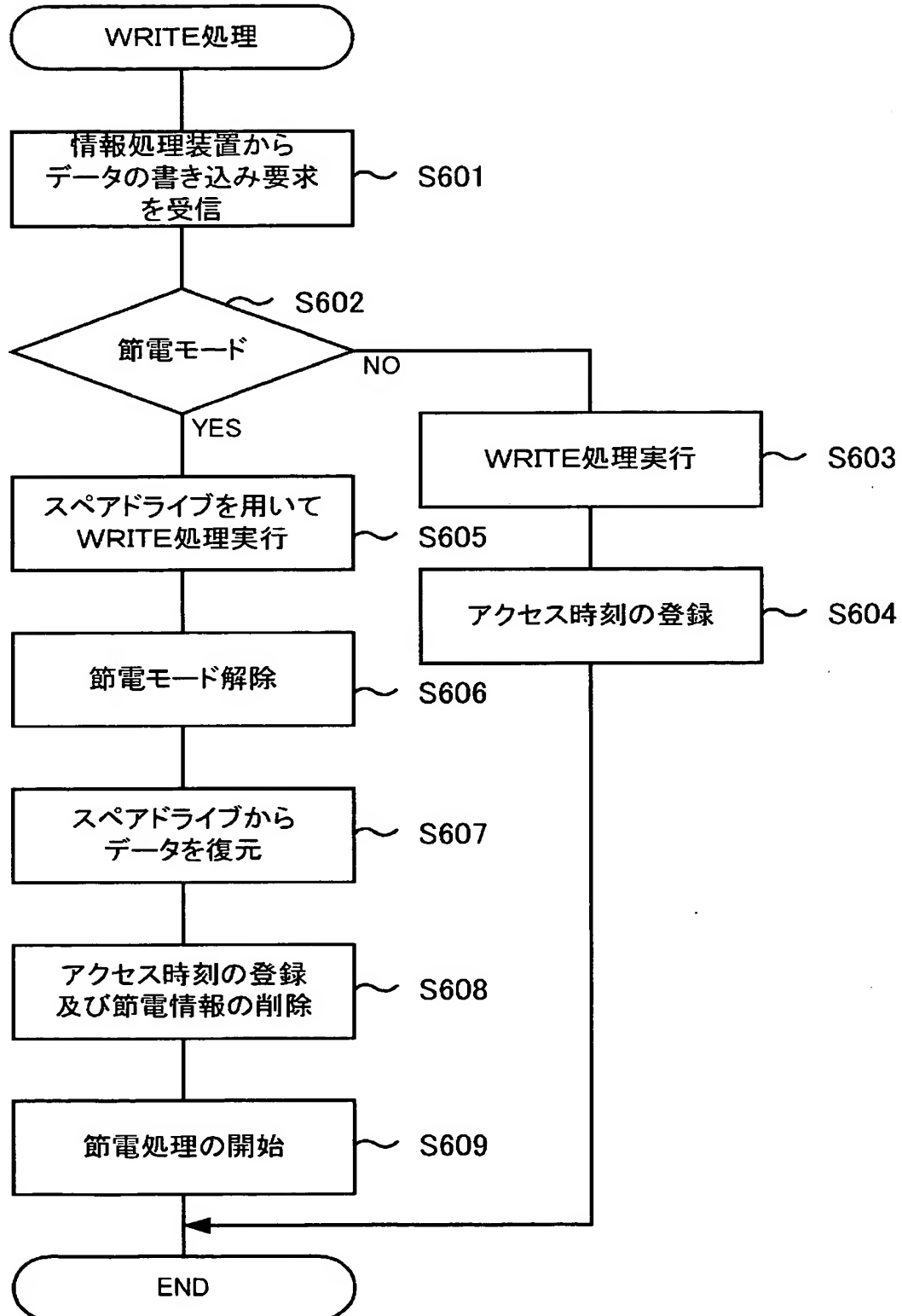
【図 4】



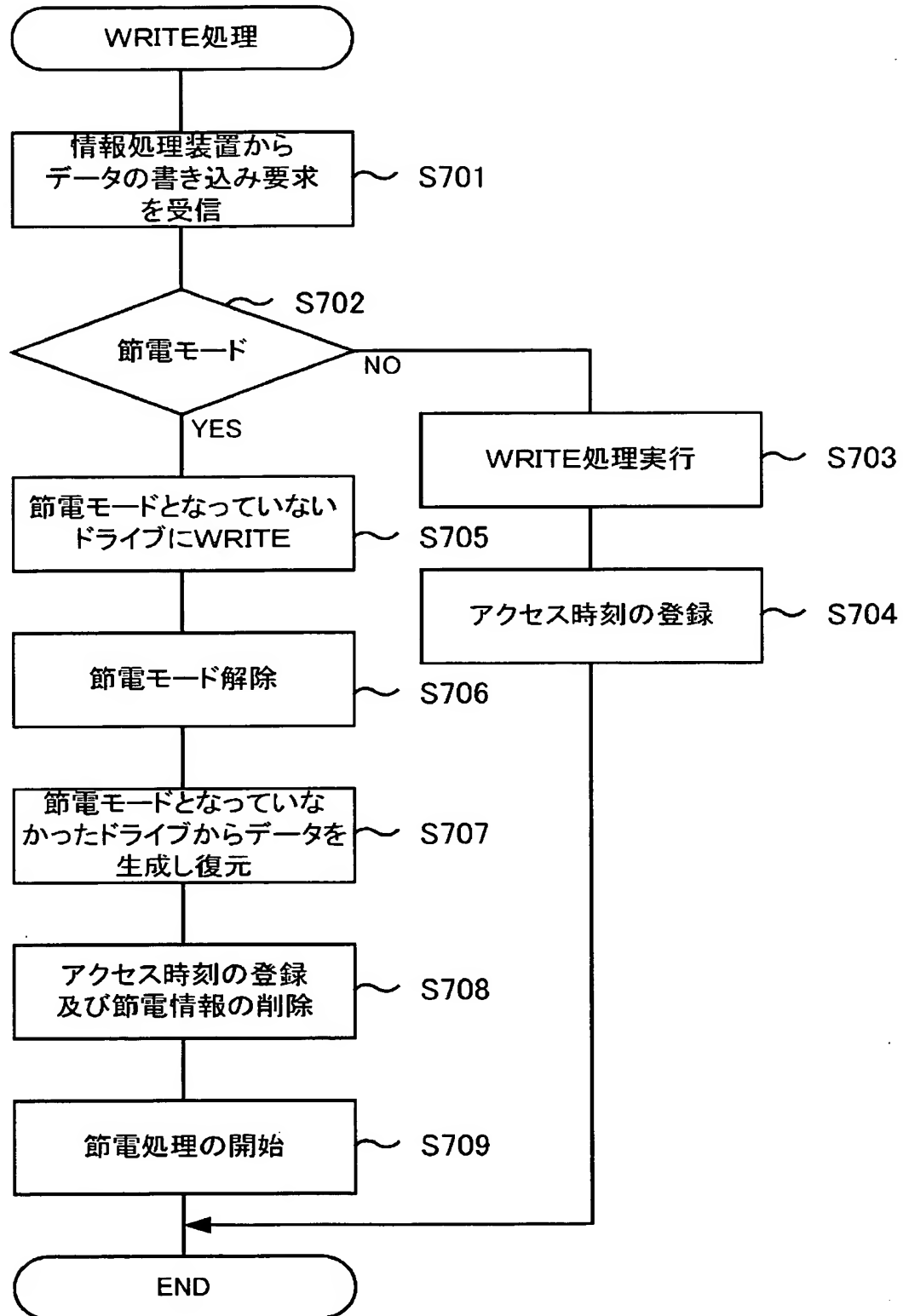
【図 5】



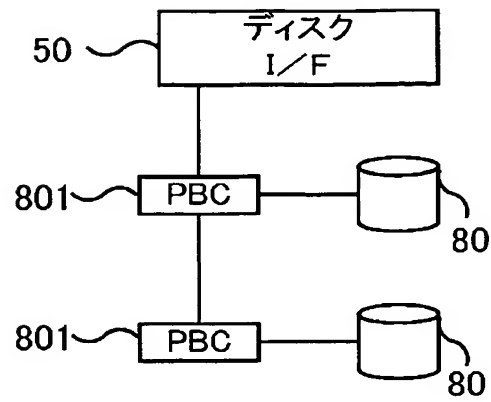
【図 6】



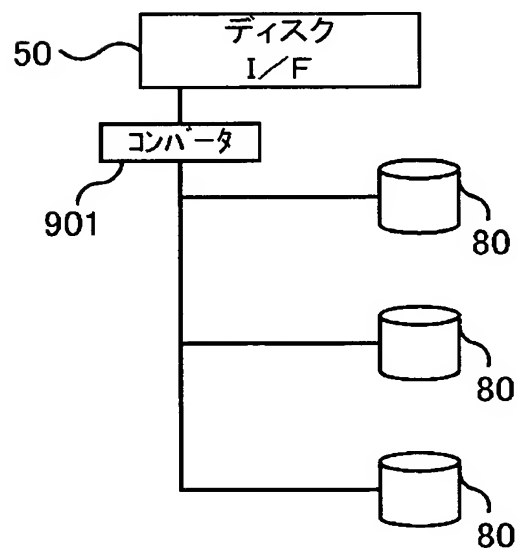
【図 7】



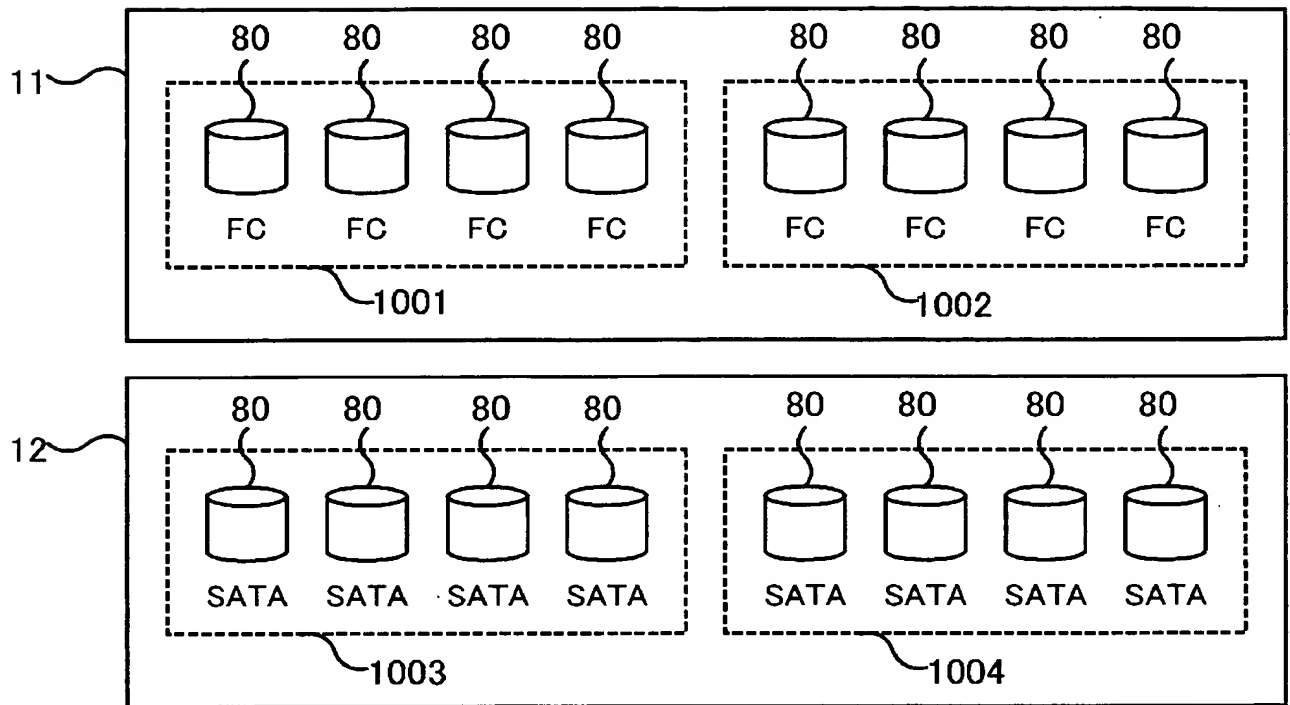
【図 8】



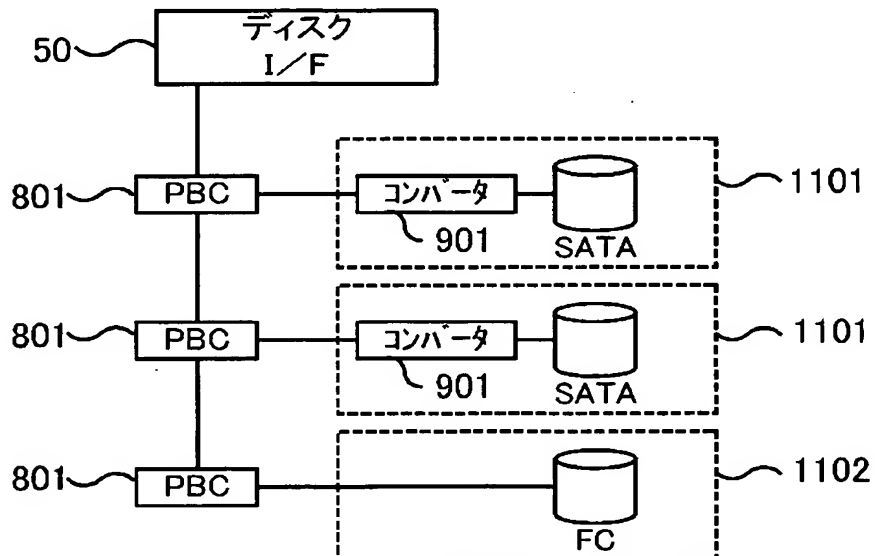
【図 9】



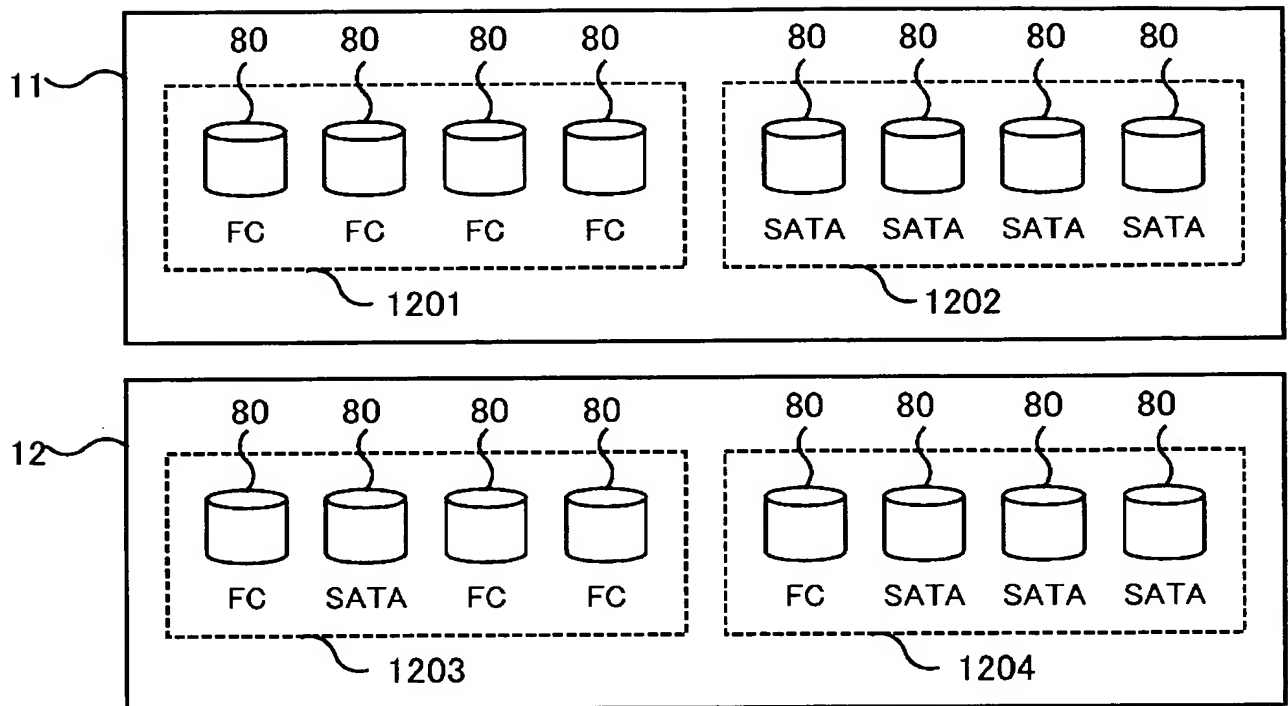
【図 10】



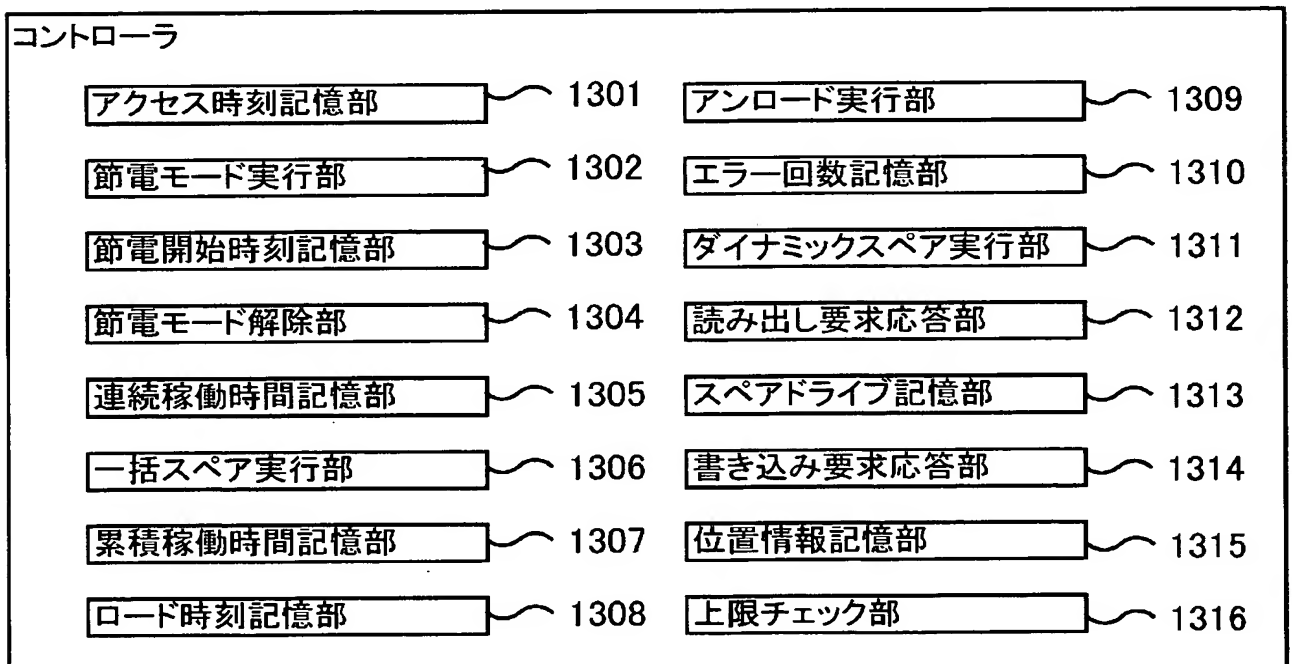
【図 11】



【図 1 2】



【図 1 3】



13

【図 14】

メモリ

設定 RAID	管理 LUN	ドライブ タイプ	利用 形態	アクセス 時刻	節電開始 時刻	連続 稼働時間
1	1, 2	FC		12:15		30分
2	3	SATA		11:40	12:00	0
3	4, 5	混在	ニアライン	10:30	10:50	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮

1401

設定 RAID	ドライブ 番号	ロード 時刻	累積 稼働時間	アンロード 回数	エラー 回数
1	1	11:00	500	50	3
1	2	11:10	350	120	5
⋮	⋮	⋮	⋮	⋮	⋮

1402

節電待ち時間	節電時間
20分	60分

1403

スベア待ち時間	スベア時間
60分	60分

1404

アンロード待ち時間
200分

1405

上限稼働時間	上限アンロード回数
50,000時間	100,000回

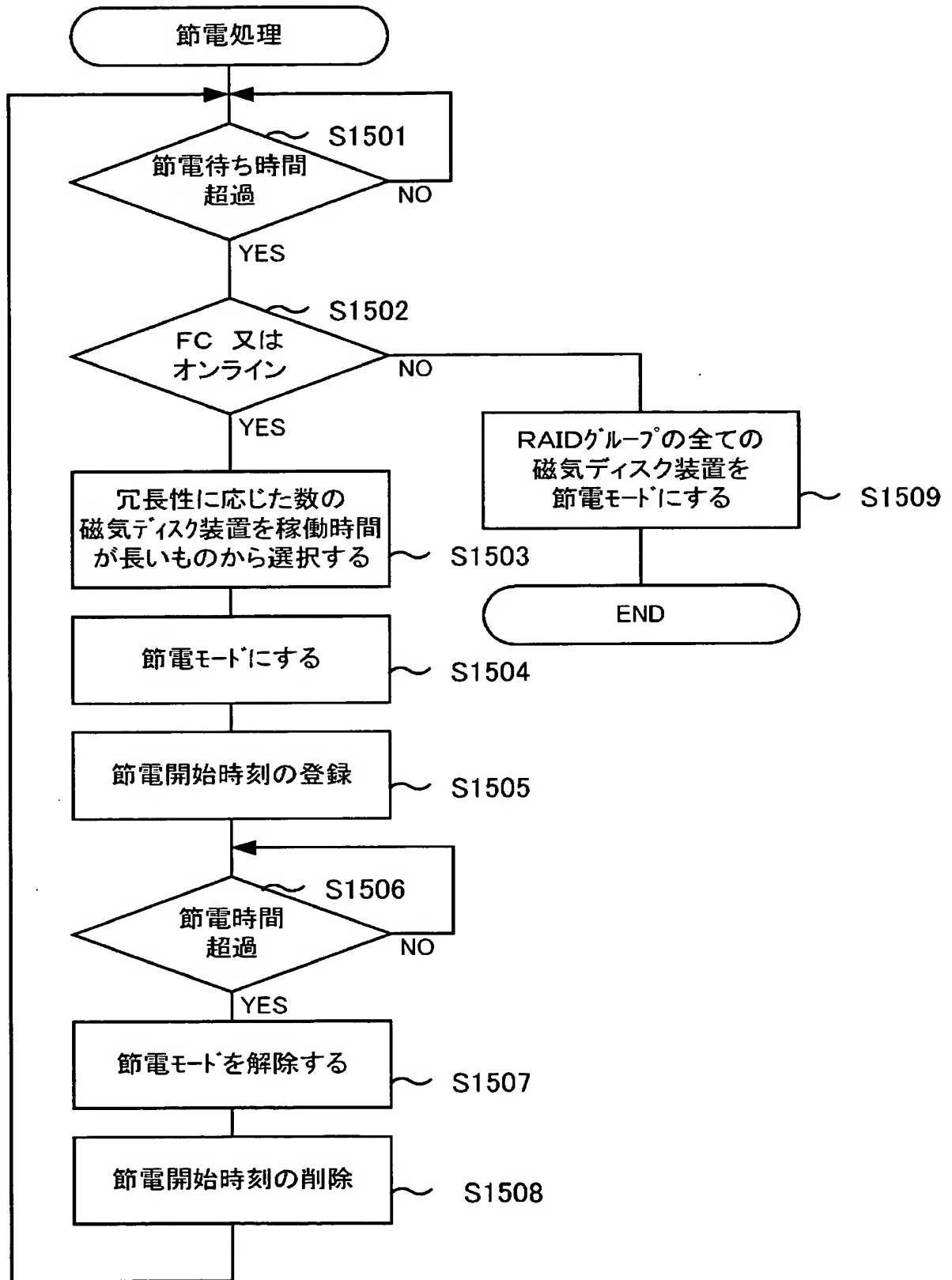
1406

稼働時間	上限エラー回数
0~1000	10
1001~2000	5
⋮	⋮

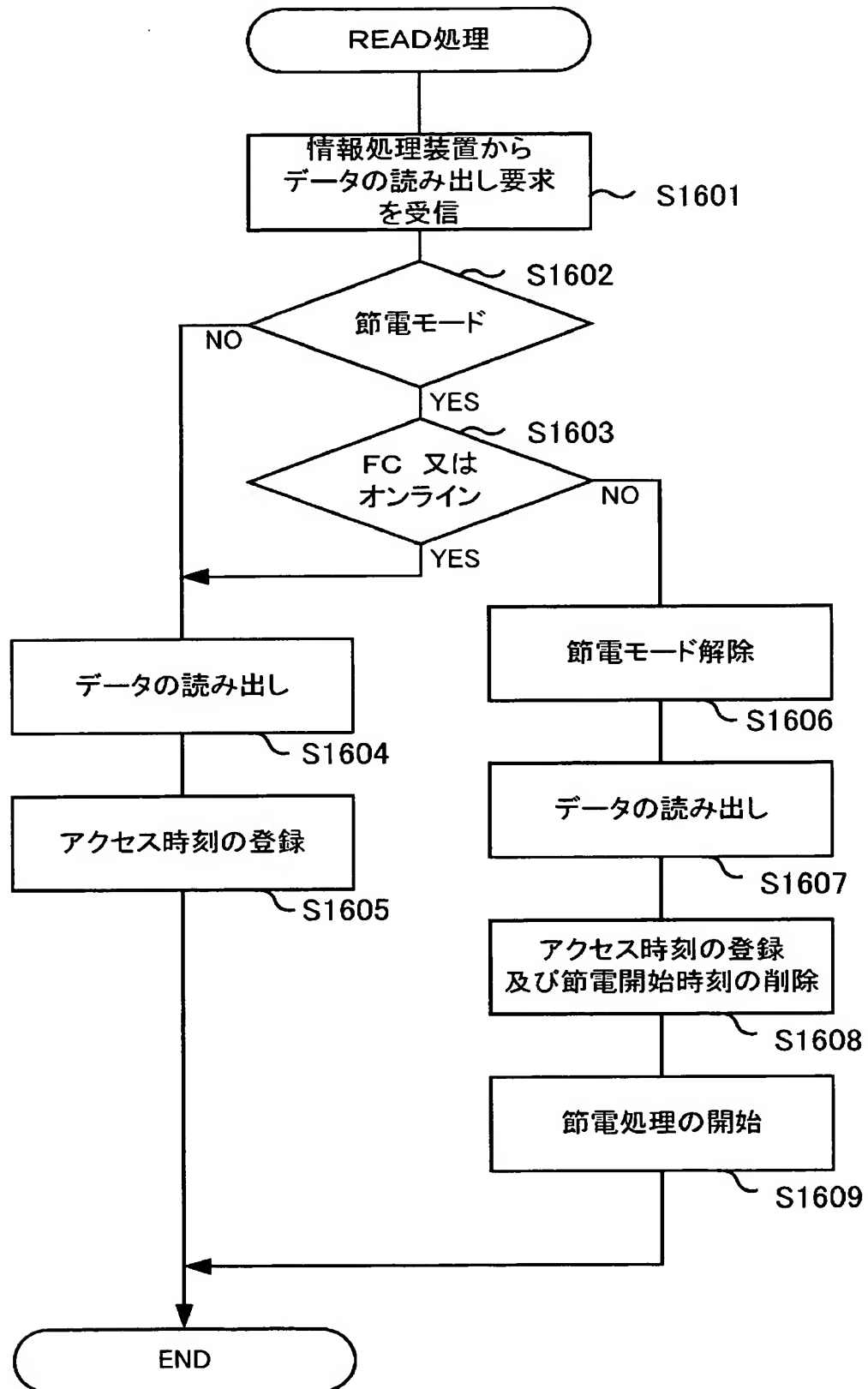
1407

15

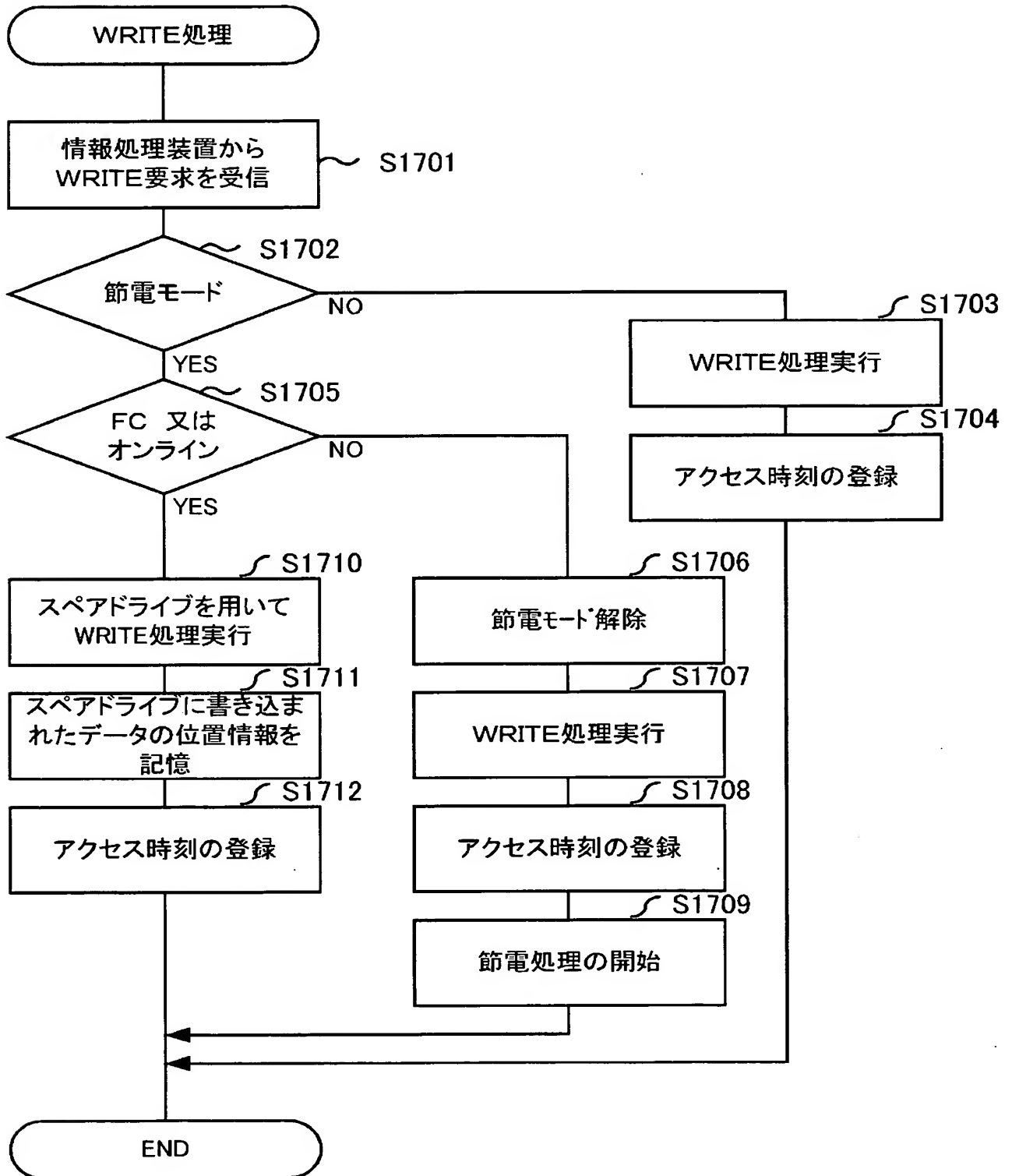
【図 15】



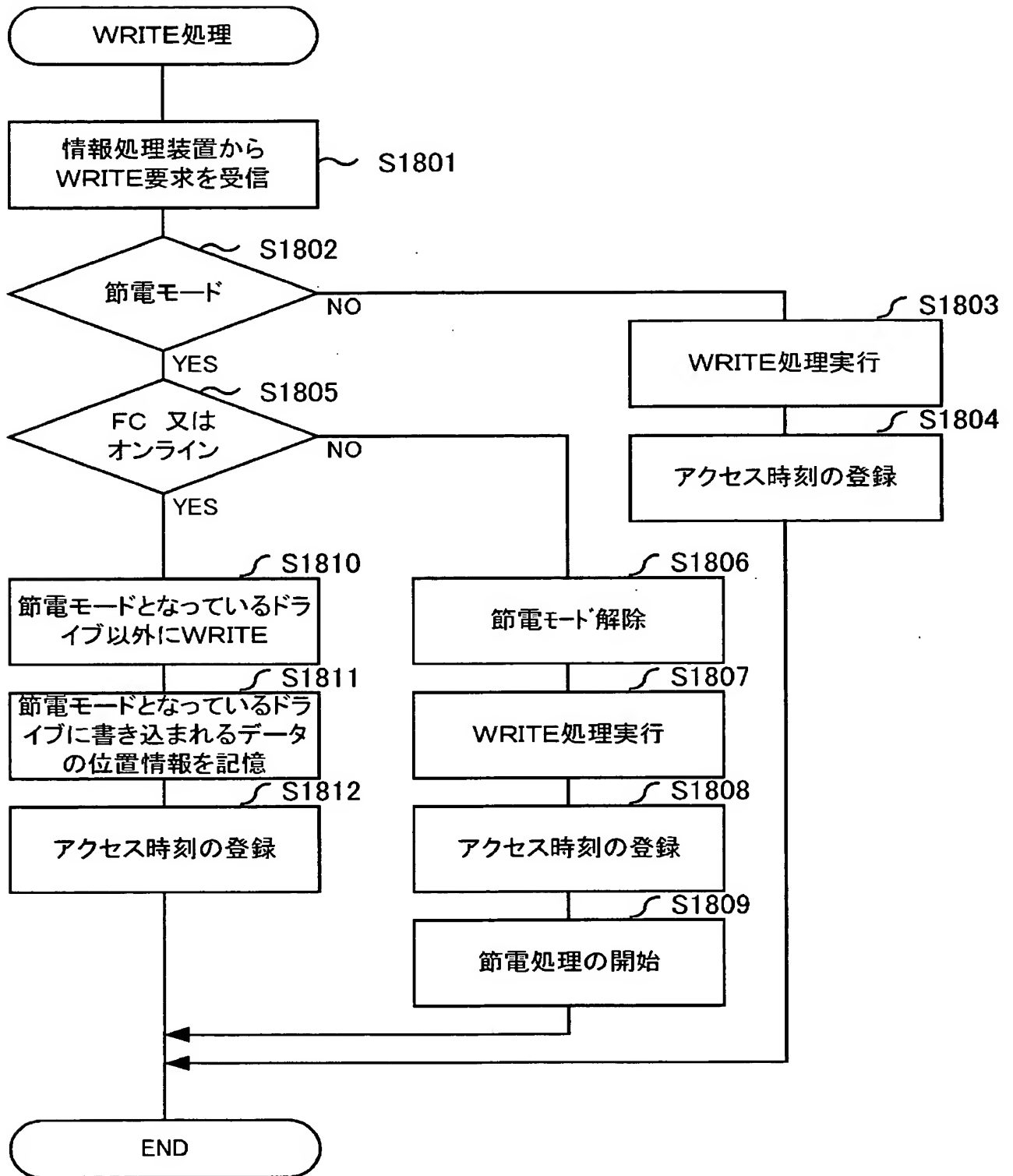
【図 16】



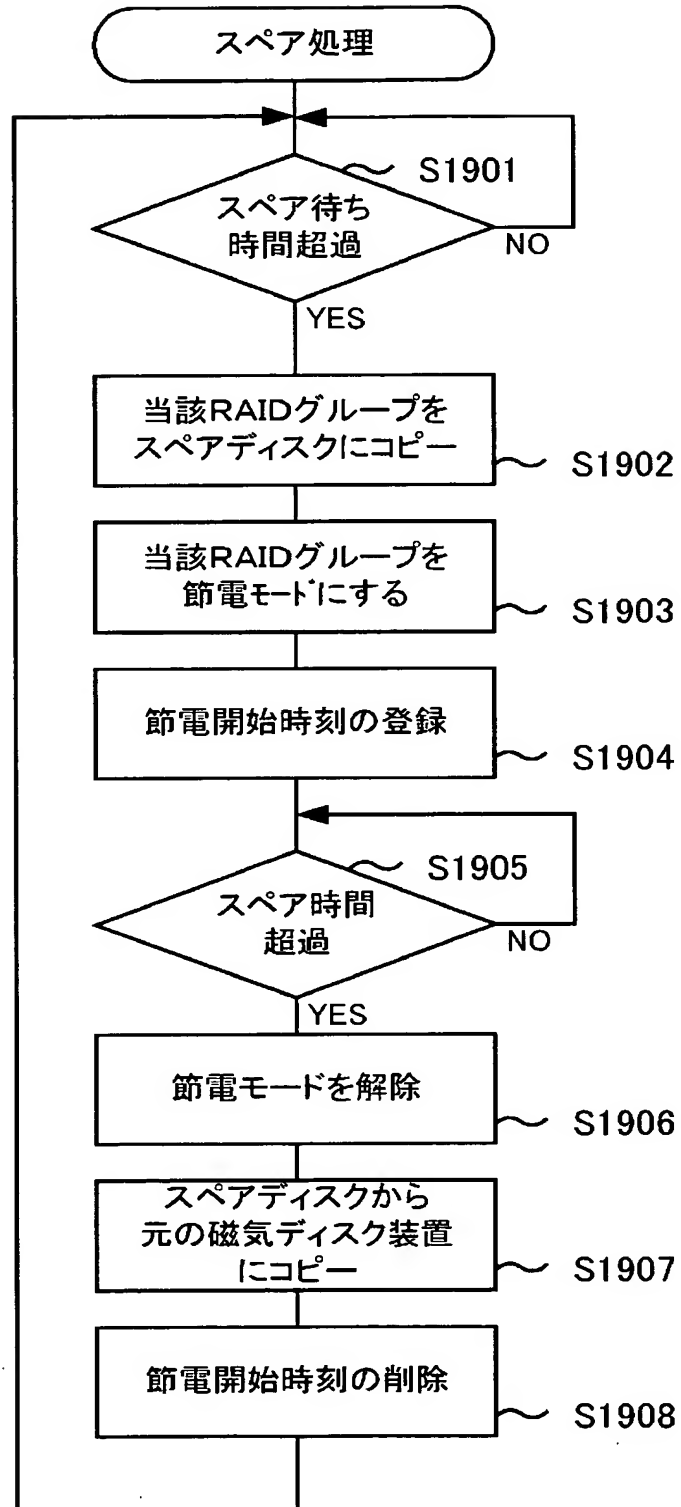
【図 17】



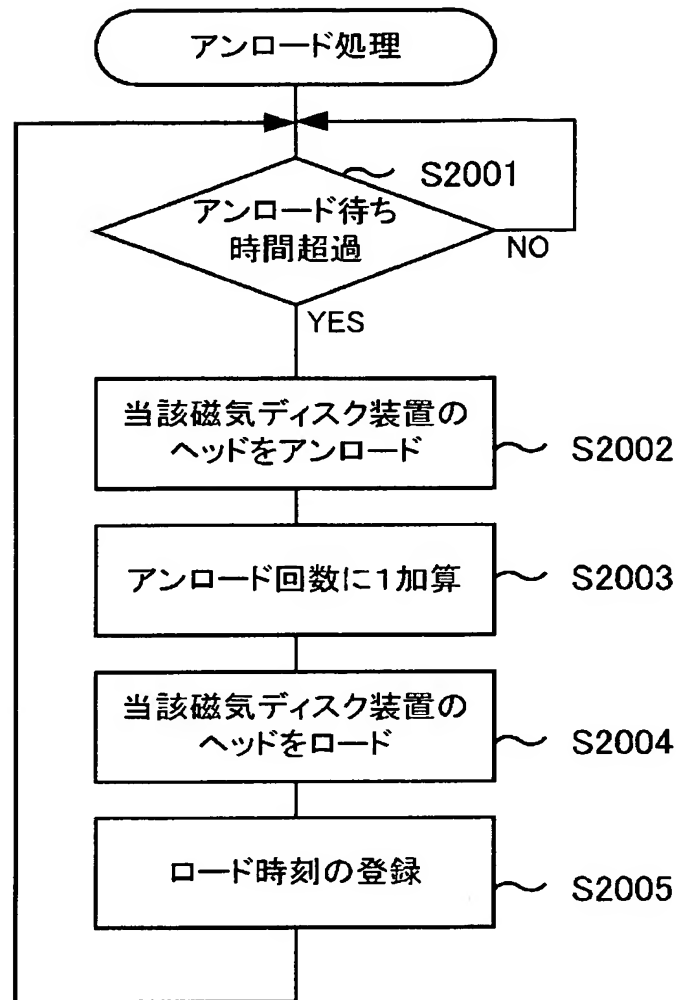
【図 18】



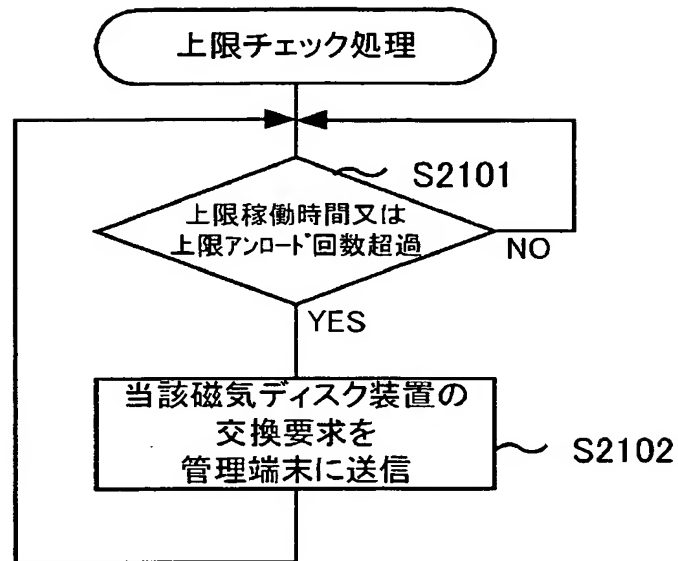
【図 19】



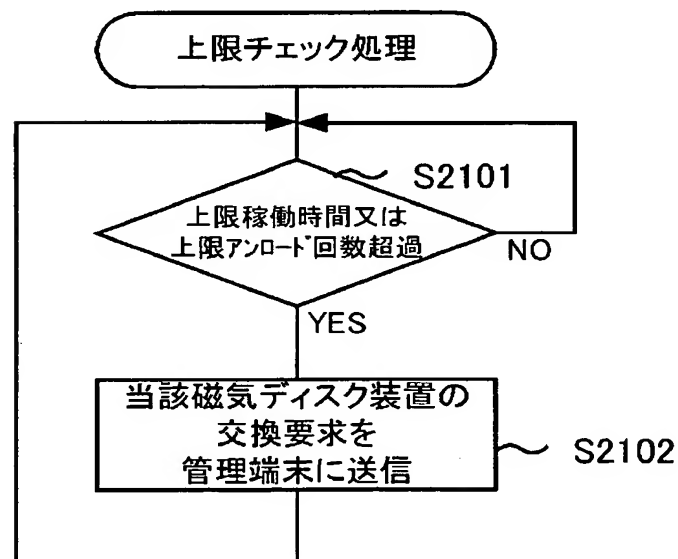
【図 20】



【図 21】



【図 22】



【書類名】 要約書**【要約】**

【解決手段】 論理ボリュームに対するデータの読み出し要求又は書き込み要求を受信した時刻をメモリに当該論理ボリュームを構成する R A I D グループのアクセス時刻として記憶し、アクセス時刻が所定の時間を経過している場合、R A I D グループの冗長性に応じた数の前記ハードディスクドライブを節電モードとする。また、第一のインタフェース規格の第一のハードディスクドライブと、第一のハードディスクドライブよりも寿命の短い、第二のインタフェース規格の第二のハードディスクドライブとを有し、アクセス時刻が所定の時間を経過している R A I D グループについて、R A I D グループが第一のハードディスクドライブのみである場合は、R A I D グループの冗長性に応じた数の第一のハードディスクドライブを節電モードとし、R A I D グループが第二のハードディスクドライブのみである場合は、任意の数の第二のハードディスクドライブを節電モードとする。

【選択図】 図 1

特願 2 0 0 4 - 0 0 9 0 0 4

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 5 1 0 8]

1. 変更年月日

1 9 9 0 年 8 月 3 1 日

[変更理由]

新規登録

住 所

東京都千代田区神田駿河台 4 丁目 6 番地

氏 名

株式会社日立製作所